SIAM J. Discrete Math.
Vol. 25, No. 2, pp. 956–988

DECAY OF CORRELATIONS FOR SPARSE GRAPH ERROR CORRECTING CODES*

SHRINIVAS KUDEKAR† and NICOLAS MACRIS‡

Abstract. The subject of this paper is transmission over a general class of binary-input memoryless symmetric channels using error correcting codes based on sparse graphs, namely, low-density generator-matrix and low-density parity-check codes. The optimal (or ideal) decoder based on the posterior measure over the code-bits and its relationship to the suboptimal belief propagation decoder are investigated. We consider the correlation (or covariance) between two code-bits, averaged over the noise realizations, as a function of the graph distance for the optimal decoder. Our main result is that this correlation decays exponentially fast for given low-density generator-matrix codes and a high enough noise parameter and also for given low-density parity-check codes and a low enough noise parameter. This has many consequences. Appropriate performance curves—called generalized extrinsic information transfer (GEXIT) functions—of the belief propagation and optimal decoders match in high/low noise regimes. This means that in high/low noise regimes the performance curves of the optimal decoder can be computed by density evolution. Another interpretation is that the replica predictions of spin-glass theory are exact. Our methods are rather general and use cluster expansions first developed in the context of mathematical statistical mechanics.

Key words. graphical codes, decay of correlations, belief propagation, cluster expansions, channel communication

AMS subject classifications. 94B05, 82B20, 82B44

DOI. 10.1137/090751827

1. Introduction.

Low-density parity-check (LDPC) codes based on sparse graphs have emerged as a focal point in the theory of error correcting codes used in noisy channel communication, largely because they are amenable to low complexity decoding and at the same time have a good performance (measured as the gap to Shannon's capacity). An important class of low complexity decoders are the message passing iterative decoders. In this framework, in order to decode a bit attached to a node of the graph, one unravels a computational tree (or covering tree) and iteratively updates messages (suitable functions of the channel output observations) passed along the edges of the computational tree. We refer to the recent book [1] for the state-of-the-art of this general theory. One would also like to be able to compare suboptimal message passing decoders with the optimal or ideal decoder. The latter is based on the posterior probability distribution supported on code-bits and is optimal in the sense that it is known to minimize the bit-error-rate among all decoders (it is also called a maximum a posteriori probability (MAP) decoder, and this is the terminology that we adopt in this paper). A priori the comparison of decoders is not easily done since the MAP decoder is in general computationally complex.

One of the most important low complexity message passing decoders is the belief propagation (BP) algorithm. It is well known that for a code whose graph is a tree, the BP algorithm has the same performance as the MAP decoder. This essentially comes

*Received by the editors March 9, 2009; accepted for publication (in revised form) November 12, 2010; published electronically July 1, 2011.

http://www.siam.org/journals/sidma/25-2/75182.html

†New Mexico Consortium and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM (skudekar@lanl.gov). This author's work has been supported by a grant of the Swiss National Foundation 200020-113412.

‡Communication Theory Laboratory, School of Computer and Communication Sciences, Ecole Polytechnique Fédérale Lausanne, CH-1015 Lausanne, Switzerland (nicolas.macris@epfl.ch).

from the fact that on a tree the computational graph of a node matches the original graph itself. However, codes based on tree graphs have poor performance, and one needs to consider graphs with loops or cycles. With cycles in the original graph, the messages on the computational tree are no longer independent, and it is not a priori clear if and why the BP algorithm should retain any close relationship to the MAP decoder. A fundamental theoretical tool that allows us to analyze the BP algorithm is density evolution (DE) first developed in [2]. From DE one can, for example, obtain a noise threshold above which reliable communication is not possible with BP decoding. The analysis proceeds by taking first a very large block length n and looking at $d \ll n$ iterations of the BP decoder. Eventually one considers the asymptotics $\lim_{d\to+\infty}\lim_{n\to+\infty}$. However, in the practical use of the decoder one fixes n (large), and the iterations $d \gg n$ are performed until one reaches an acceptably small bit-error-rate. This corresponds to the asymptotics $\lim_{n\to+\infty} \limsup/\inf_{d\to+\infty}$. The practical success of density evolution relies on the equivalence of these two limiting procedures, an open problem in general.

It is fair to say that these issues have been resolved over the binary erasure channel (BEC) [1] (the analysis, however, has not proceeded from the point of view of the correlations of the MAP decoder). An important tool in the analysis over the BEC has been the "extended BP" extrinsic information transfer (EXIT) curve (which is a suitable continuation of the bit-error-rate curve under BP decoding). Recently it was shown that the bit-error-rate of the MAP decoder can be obtained from that of the extended BP EXIT curve by a Maxwell construction, just as in the theory of first order phase transitions[1] [6], [7]. This construction allows us to compute the MAP noise threshold and to compare it to the BP noise threshold. The validity of the exchange of limits $d \to +\infty$ and $n \to +\infty$ (for the BP decoder) can also be derived for the BEC using natural monotonicity properties of the decoder [1].

For the case of transmission over more general channels very little is known about these issues. Indeed there one lacks the combinatorial methods available for the BEC, and radically new methods have to be used. Convenient measures of the performance, which generalize the EXIT curves, are the so-called generalized extrinsic information transfer (GEXIT) curves [1] (see the next section for their precise definition). It is believed that in terms of these, the results obtained for the BEC still hold. In particular, the GEXIT curves for the BP and the MAP decoder should match for high and low noise regimes away from the phase transition thresholds. Such conjectures are supported by spin glass theory calculations (e.g., the replica and cavity methods), which provide conjectural but analytic formulas. One-sided bounds have been derived for the GEXIT curves by the (information theoretical) method of physical degradation [1] and also by using correlation inequalities valid for spin glasses [13]. Related bounds on the conditional input-output entropy have also been derived [3], [4] by using "interpolation methods" first developed in the mathematical theory of spin glasses [5], [8], [9]. As it turns out, all these bounds match the replica expressions and are therefore believed to be the best possible. In [10] the interpolation method has been extended to obtain the converse bounds for a class of Poissonian LDPC codes over the BEC, thus recovering combinatorial results of [11] in a completely different way. Concerning the problem of exchanging the $d, n \to +\infty$ limits, we refer to [15] for recent progress that goes beyond the BEC.

[1]The extended BP curve corresponds to the pressure-volume curve of the Van der Waals theory of the liquid-gas transition, and the MAP curve corresponds to the isotherms obtained by Maxwell's equal area construction.

In this work we will show that a good deal can be learned by looking at the correlations (more precisely the covariance), averaged over the channel outputs of the MAP decoder. We comment below about the methods used, but let us say at the outset that our aim is to cover a fairly general class of binary input memoryless symmetric channels, including the binary symmetric and Gaussian ones. One of our main results is that for sufficiently low noise (LDPC codes) the correlations between two code-bits decay exponentially fast as a function of the graph distance between the two code-bits, uniformly in the block length size n. The sparsity of the underlying graph then implies that if furthermore the decay rate beats the local expansion of the graph, the MAP GEXIT curve can be computed by DE. Another interpretation of this result is that the solutions provided by the replica/cavity methods of spin glass theory are exact.

Low-density generator-matrix codes (LDGM) codes have a very clear relationship to spin glass models on random graphs, and it is useful to study them before we can attack the harder case of LDPC codes. Besides, the present analysis could potentially be useful in other contexts where they are used (e.g., rateless codes, source coding). For high noise we prove the decay of correlations and that the MAP GEXIT curve can be computed by DE. For that system, we can also show that the decay of correlations implies that the limits $d, n \to +\infty$ can be exchanged for the BP decoder, at least on the binary symmetric channel.

Let us say a few words about the techniques used in this work. The study of the behavior of correlations as a function of the distance between local degrees of freedom is one of the central aims of statistical physics. For lattice spin systems (e.g., the Ising model) an important criterion that ensures correlation decay is Dobrushin's criterion [12]—which is of probabilistic nature—and its various improvements. The main other method—which is not necessarily of probabilistic nature—is based on suitable expansions in powers of "the strength of interactions." There exists a host of such expansions collectively called "cluster expansions," and in the context of spin systems the first and simplest such expansion is the so-called polymer expansion [16]. The main rule of thumb is that all these methods work if the degrees of freedom are weakly interacting or if one can transform the original system into an effective one involving new weakly interacting degrees of freedom. It turns out that sophisticated forms of the cluster expansions can be carried out, for LDGM codes in a high noise regime and for LDPC codes, in a low noise regime, for a fairly general class of channels including the binary symmetric channel (BSC) as well as the binary input additive white Gaussian noise channel (BIAWGNC). As we will explain later, it is necessary to use quite sophisticated cluster expansions for at least two reasons. Concerning LDGM codes, Dobrushin's criterion and the polymer expansion require bounded channel outputs (and thus do not cover the case of the BIAWGNC). Concerning LDPC codes, one has to transform the system to a dual one that involves "negative Gibbs weights" and cannot even be treated by probabilistic methods.

The rest of the paper is organized as follows. In section 2 we formulate the models and give a unified view of the main results both for LDGM and LDPC codes. The main strategy of the proofs is also explained there. Sections 3 and 4 contain the proofs of correlation decay and its consequences for the GEXIT curves. The problem of exchanging the limits of an iteration number and block length size are addressed in section 5. We conclude by pointing out open problems and further connections to the recent literature. The appendix reviews in a streamlined form the two cluster expansions that are used in sections 3 and 4.

Summaries of the present results have been reported for the special case of the BIAWGNC [17], [18].

2. Models and main results. We consider binary-input memoryless output-symmetric channels defined by a transition probability distribution function $p_{Y|X}(y|x)$ with inputs $x \in \{-1,+1\}$ and outputs belonging to $\mathbb{R}$. Since we use techniques from statistical mechanics, it is convenient to immediately map the usual input alphabet $\{0,1\}$ to $\{-1,+1\}$. Symmetry of the channel means that $p_{Y|X}(-y|-x) = p_{Y|X}(y|x)$. The intensity of the noise is called ϵ. It will be convenient to trade off the channel outputs y for the half–log-likelihood

$$l = \frac{1}{2}\ln\left[\frac{p_{Y|X}(y|+1)}{p_{Y|X}(y|-1)}\right]. \tag{1}$$

It is well known that on a symmetric channel one can assume without loss of generality that the all-one codeword (i.e., the usual all-zero codeword) is transmitted, and therefore the channel outputs are independently and identically distributed (i.i.d.) with distribution $p_{Y|X}(y|+1)dy \equiv c(l)dl$. Throughout the paper $c(l)$ need not be a density; all formulas are valid as long as it is interpreted as a distribution. For clarity, we will assume that the noise parameter varies in an interval $[0,\epsilon_{\max}]$ ($\epsilon_{\max}$ is possibly infinite), where $\epsilon \to 0$ corresponds to low noise and $\epsilon \to \epsilon_{\max}$ corresponds to high noise. The general class of channels for which our main results hold is as follows.

DEFINITION 1 (class of channels). *We define the class $\mathcal{K}$ of binary-input memoryless output-symmetric channels as follows:*

1. *The numbers $T_{2p}(\epsilon) = \frac{d}{d\epsilon}\int_{-\infty}^{\infty} dl c(l)(\tanh l)^{2p}$ are bounded uniformly with respect to a $p \geq 1$ integer.*
2. *For any finite $m > 0$ we have $\mathbb{E}[e^{m|l|}] \leq c_m < +\infty$.*
3. *(Low noise condition.) There exists $s_0 > 0$ small enough such that for $0 < s \leq s_0$ we have $\lim_{\epsilon\to 0}\mathbb{E}[e^{-sl}] = 0$.*
4. *(High noise condition.) Set $\delta(\epsilon,H) = e^{4H} - 1 + \mathbb{P}(|l| > H)$. One can find $H(\epsilon)$ such that $\lim_{\epsilon\to\epsilon_{\max}}\delta(\epsilon,H(\epsilon)) = 0$.*

Note that this class is not the most general that we can treat, but it is at the same time fairly general and keeps the analysis at a technically reasonable level. Let us give some important explicit examples.

BSC. The BSC channel has binary inputs $x \in \{-1,+1\}$. The output is given by a flip $\pm 1 \to \mp 1$ occuring with probability ϵ. The transition probability and half–log-likelihood distributions are

$$\begin{aligned} p_{Y|X}(y|x) &= (1-\epsilon)\delta(y-x) + \epsilon\delta(y+x), \\ c(l) &= (1-\epsilon)\delta\left(l - \frac{1}{2}\ln\frac{1-\epsilon}{\epsilon}\right) + \epsilon\delta\left(l - \frac{1}{2}\ln\frac{\epsilon}{1-\epsilon}\right). \end{aligned} \tag{2}$$

Clearly $\epsilon = \frac{1}{2}$ corresponds to "infinite noise." Thus we will keep $0 \leq \epsilon\epsilon_{\max} =\leq \frac{1}{2}$ (the case $\frac{1}{2} \leq \epsilon \leq 1$ being symmetric). One can check that the conditions of class $\mathcal{K}$ are met with $T_{2p}(\epsilon) = 2p(1-2\epsilon)^{2p-1}$, $\mathbb{E}[e^{m|l|}] = (\frac{1-\epsilon}{\epsilon})^{m/2}$, $\mathbb{E}[e^{-sl}] = \epsilon^{s/2}(1-\epsilon)^{1-(s/2)} + (1-\epsilon)^{s/2}\epsilon^{1-(s/2)}$, and $H(\epsilon) = \log\frac{1-\epsilon}{\epsilon}$.

BIAWGNC. The BIAWGNC has binary inputs $x \in \{-1,+1\}$ and output $y = x + \sqrt{\epsilon}\xi$ with ξ the standard Gaussian random variable. Here ϵ is the noise variance to be interpreted as the inverse of the "signal to noise" ratio. The transition probability and half–log-likelihood distributions are

$$p_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\epsilon}} \exp\left(-\frac{(y-x)^2}{2\epsilon}\right), \qquad c(l) = \frac{\epsilon}{\sqrt{2\pi}} \exp\left(-\frac{(l-\epsilon^{-1})^2}{2\epsilon^{-1}}\right). \tag{3}$$

Here $\epsilon_{\max} = +\infty$. Again one can check that the conditions are met with $T_{2p}(\epsilon) \leq \int_{-\infty}^{+\infty} dl |\frac{dc(l)}{d\epsilon}|$, $\mathbb{E}[e^{m|l|}] < \infty$, $\mathbb{E}[e^{-sl}] = e^{-s\epsilon^{-1}(1-(s/2))}$, and $H(\epsilon) = 2\epsilon^{-1/4}$.

BEC. The BEC has inputs $x \in \{-1, +1\}$ and outputs $y \in \{-1, E, +1\}$. Here E is an "erasure" symbol. Formally $p_{Y|X}(y|x) = (1-\epsilon)\delta(y-x) + \epsilon\delta(y-E)$ (here one may consistently set $E = 0$) and $c(l) = \epsilon\delta(l) + (1-\epsilon)\delta_{+\infty}(l)$. Note that the BEC is not contained in the class $\mathcal{K}$ because of the second condition. Nevertheless due to the special nature of this channel our methods can easily be adapted, but we will not give the details here since this is a case that has already been thoroughly analyzed in the literature [1].

Given LDGM codes are constructed from a given bipartite graph with m information-bit nodes (variable nodes) and n code-bit nodes (check nodes), and edges connecting variable and check nodes only. The design rate of the code $R = \frac{m}{n}$ is kept fixed. The set of neighbors of a variable node a is called ∂a, and the set of neighbors of a check node i is called ∂i. We consider graphs with bounded node degrees $|\partial a| \leq l_{\max}$ and $|\partial i| \leq k_{\max}$. Information bits $u_1, \ldots, u_m \in \{-1, +1\}^m$ are attached to the variable nodes, and the code-bits $x_1, \ldots, x_n$ attached to the check nodes are obtained as

$$x_i = \prod_{a \in \partial i} u_a, \qquad i = 1, \ldots, n. \tag{4}$$

We also consider ensembles of such codes defined by random graph constructions. We do not explain the details of these constructions here, except for saying that an LDGM(Λ, P) ensemble is specified by the generating functions of variable (resp., check) node degree distributions $\Lambda(z) = \sum_{l=1}^{l_{\max}} \Lambda_l z^l$ (resp., $P(z) = \sum_{r=1}^{r_{\max}} P_r z^r$) [1].

Given that LDPC codes are similarly constructed from a given bipartite graph with n variable nodes $i = 1, \ldots, n$ (this time these are the code-bit nodes) and m check nodes $c = 1, \ldots, m$, with edges connecting variable and check nodes only, the design rate $R = 1 - \frac{m}{n}$ is fixed. We assume that the node degrees are bounded $|\partial i| \leq l_{\max}$ and $|\partial c| \leq k_{\max}$. The code-bits $x_1, \ldots, x_n$ attached to the variable nodes satisfy m parity check constraints

$$\prod_{i \in \partial c} x_i = 1, \qquad c = 1, \ldots, m. \tag{5}$$

We also consider ensembles of such codes defined by random graph constructions; an ensemble is specified by the generating functions of variable (resp., check) node degree distribution $\Lambda(z) = \sum_{l=1}^{l_{\max}} \Lambda_l z^l$ (resp., $P(z) = \sum_{r=1}^{r_{\max}} P_r z^r$) [1].

The optimal MAP decoder is based on the posterior measure of the transmitted codeword given the received message $y^n = (y_1, \ldots, y_n)$. For LDGM codes this conditional measure is best viewed as being supported on information bits,

$$p_{U^m|Y^n}(u^m|y^n) = \frac{1}{Z} \prod_{i=1}^{n} e^{l_i \prod_{a \in \partial i} u_a}. \tag{6}$$

For LDPC codes the conditional measure is

$$p_{X^n|Y^n}(x^n|y^n) = \frac{1}{Z}\prod_{c=1}^{m}\frac{1}{2}\left(1+\prod_{i\in\partial c}x_i\right)\prod_{i=1}^{n}e^{l_i x_i}. \tag{7}$$

In both cases Z is the appropriate normalization factor. These measures are random because of the channel outputs and possibly because the code is chosen at random from an ensemble. The average with respect to the channel outputs is often denoted by $\mathbb{E}_{l^n}$, and the average with respect to a code ensemble is generically denoted by $\mathbb{E}_{\mathcal{C}}$. We will also use the notation $\mathbb{E}_{l^{n\setminus i}}$ when the average is over all outputs except the ith one. A crucial point is that the interactions or constraints in these measures are local so that they can be analyzed with the tools developed in the theory of Gibbs measures [12]. We use the bracket notation

$$\langle f\rangle = \sum_{u^m} f(u^m)p_{U^m|Y^n}(u^m|y^n) \quad \text{or} \quad \langle f\rangle = \sum_{x^n} f(x^n)p_{X^n|Y^n}(x^n|y^n) \tag{8}$$

for the Gibbs averages of functions f. The weight with respect to which the average is calculated will be clear from the context. It turns out that even for (6) we will only need to look at averages of functions of the transmitted code-bits x^n; for example, $\langle x_i\rangle = \langle\prod_{i\in a}u_a\rangle$. It is important to remember that the bracket is defined for finite n, although we do not write explicitly $\langle -\rangle_n$ to alleviate the notations.

Let us now define the performance measures known as the GEXIT function. The formalism adopted here follows [13]. A more informational theoretic point of view as well as their use in coding theory may be found in [1]. The average (over noise realizations) Gibbs entropy of the two measures is nothing other than Shannon's input-output conditional entropy $\frac{1}{n}H(U^m|Y^n)$, $\frac{1}{n}H(X^n|Y^n)$, denoted in both cases by h_n. The MAP-GEXIT function is simply defined as the ϵ derivative of this conditional entropy. We define

$$g_n(\epsilon) = \frac{d}{d\epsilon}\mathbb{E}_{\mathcal{C}}[h_n], \tag{9}$$

where the expectation $\mathbb{E}_{\mathcal{C}}$ is over an LDGM or LDPC code ensemble. When this derivative is performed, one finds that the MAP-GEXIT function is a functional of the soft-bit MAP estimate (the magnetization) $\langle x_i\rangle$. It is much more convenient, in fact, to express it as a functional of the extrinsic estimate $\langle x_i\rangle_0$, which is the Gibbs average computed for $l_i = 0$. We have (see, for example, [13] for an explicit derivation)

$$g_n(\epsilon) = \mathbb{E}_{\mathcal{C}}[\mathcal{G}(\langle x_i\rangle_0)], \tag{10}$$

where in the case of LDGM codes

$$\mathcal{G}(\langle x_i\rangle_0) = \frac{\Lambda'(1)}{P'(1)}\int dl_i\frac{dc(l_i)}{d\epsilon}\mathbb{E}_{l^{n\setminus i}}\ln\left\{\frac{1+\langle x_i\rangle_0\tanh\ l_i}{1+\tanh\ l_i}\right\} \tag{11}$$

while for LDPC codes

$$\mathcal{G}(\langle x_i\rangle_0) = \int dl_i\frac{dc(l_i)}{d\epsilon}\mathbb{E}_{l^{n\setminus i}}\ln\left\{\frac{1+\langle x_i\rangle_0\tanh\ l_i}{1+\tanh\ l_i}\right\}. \tag{12}$$

Note that the only formal difference between the two cases is in the normalization factor, but of course the Gibbs average in each expression pertains to two different measures. A

clarification is in order concerning the derivative with respect to ϵ in these formulas. In fact differentiability of the channel distribution $c(l_i)$ is not really needed: one may allow the ith bit to be transmitted with an independent noise value ϵ_i, compute the derivative with respect to ϵ_i *outside* of the integral, and then set ϵ_i equal to ϵ. The slightly abusive notation has been adopted here to alleviate the formalism. These explicit forms of the functionals will be used in sections 3 and 4.

Let us now describe the BP decoder from the point of view of Gibbs measures. Given a graph G defining a given LDGM or LDPC code with n and m fixed (large), we choose a code-bit node i and construct the computational tree $T_d(i)$ of depth d (even). This is the universal covering tree truncated at distance d from node i. We label the variable/check nodes of this tree with new independent labels denoted n. Let $\pi : T_d(i) \to G$ be the projection from the covering tree to the original graph. A node $\nu \in T_d(i)$ has an image $\pi(\nu)$, and due to the loops in G this projection is a many-to-one map: one may have $\nu \neq \nu'$, $\pi(\nu) = \pi(\nu')$. Now, consider a tree-code defined as before (see (4), (5)) on the bipartite tree graph $T_d(i)$. One can view the BP decoder for node x_i as a MAP decoder for this tree-code. In other words the BP decoder uses the Gibbs measure on $T_d(i)$: one crucial point is that for this Gibbs measure the half–log-likelihood variables attached to the nodes are no longer independent. For the LDGM case the measure is

$$\frac{1}{Z_{T_d(i)}} \prod_{k \in T_d(i)} e^{l_\pi(k) \prod_{a \in \partial k} u_a}, \tag{13}$$

while for LDPC case

$$\frac{1}{Z_{T_d(i)}} \prod_{c \in T_d(i)} \frac{1}{2}\left(1 + \prod_{k \in \partial c} x_k\right) \prod_{k \in T_d(i)} e^{l_{\pi(k)} x_k}, \tag{14}$$

where in each case $Z_{T_d(i)}$ is the proper normalization factor. We call $\langle - \rangle_d^{BP}$ the Gibbs bracket with respect to these measures. The extrinsic BP soft-bit estimate is $\langle x_i \rangle_{0,d}^{BP}$. The BP-GEXIT function can be defined[2] in terms of the same functional as in (10):

$$g_{n,d}^{BP}(\epsilon) = \mathbb{E}_{\mathcal{C}}[\mathcal{G}(\langle x_i \rangle_{0,d}^{BP})]. \tag{15}$$

The soft-bit estimate $\langle x_i \rangle_d^{BP}$ can be computed exactly by summing the spins starting from the leaves of $T_d(i)$ all the way up to the root i. This computation is left to the reader and yields the usual message-passing BP algorithm.

We are now ready to describe our main results. The main one concerns the exponential decay of the average correlation between two code-bits x_i and x_j of a given LDGM or LDPC code, as a function of their graph distance $\mathrm{dist}(i, j)$, uniformly in the system size n.

THEOREM 1 (decay of correlations for the MAP decoder). *Consider communication over channels $\mathcal{K}$. Take an LDGM code at high enough noise $\epsilon_g < \epsilon < \epsilon_{\max}$ or an LDPC code at low enough noise $0 < \epsilon < \epsilon_p$, where $\epsilon_g, \epsilon_p > 0$ depend only on $l_{\max}, k_{\max}$. Then*

[2]The definition adopted here is very natural from the point of view of the measures (13) and (14). In [1] another definition is given that is more natural from the point of view of information theory. It is not difficult to show that they are equivalent as $n \to +\infty$.

$$\mathbb{E}_{l^n}[|\langle x_i x_j\rangle - \langle x_i\rangle\langle x_j\rangle|] \le c_1\, e^{-\mathrm{dist}(i,j)/\xi(\epsilon)}, \tag{16}$$

where c_1 is a finite positive numerical constant and $\xi(\epsilon)$ is a strictly positive constant depending only on ϵ, $l_{\max}$, and $k_{\max}$. In both regimes we have that $\xi^{-1}(\epsilon)$ grows with $\epsilon \to 0$ and $\epsilon \to \epsilon_{\max}$.

Let us say a few words on the strategy used to prove this theorem. As explained in the introduction, for LDGM at high noise and for channels with bounded log-likelihood variables, (16) follows from Dobrushin's criterion or from the polymer expansion. These, however, do not work when the likelihood variables are unbounded because, roughly speaking, overlapping polymers involve moments $\mathbb{E}[l^m]$ which can spoil the convergence as $m \to +\infty$. More physically, what happens is that even in the high noise regime there always exist with positive probability large portions of the graph that are at low noise (or "low temperature").[3] We use a very convenient cluster expansion of Dreifus, Klein, and Perez [20] that overcomes this problem by organizing the expansion over self-avoiding, random walks on the graph. Since the walks are self-avoiding, the moment problem does not occur, and we can treat unbounded loglikelihoods. For LDPC codes the situation is more subtle because of the hard parity-check constraints that give an inherently low temperature flavor to the problem. From a purely code theoretical point of view it is known that LDPC codes are the dual of LDGM codes. This algebraic duality can be exploited to transform the low noise communication model with LDPC codes to a dual model which, although not a genuine high noise communication model with LDGM codes, still retains this flavor. In fact this dual model involves "negative Gibbs weights." For this reason the cluster expansion of [20] does not work anymore, and we resort to another one first devised by Berretti [21]. The two cluster expansions have to be adapted to our setting and are therefore reviewed in a somewhat streamlined form in Appendix A.

Remark 1. The proof of Theorem 1 will make it clear that for LDGM codes on channels with bounded log-likelihoods (e.g., the BSC) at high noise, the average correlation $\mathbb{E}[\langle x_i x_k\rangle_d^{BP} - \langle x_i\rangle_d^{BP}\langle x_k\rangle_d^{BP}]$ between the root node and another one decays exponentially fast (here $\mathbb{E}$ is over the noise). See [22] for related work based on Dobrushin's criterion. The likelihood variables over $T_d(i)$ are not independent anymore so that the unbounded case is even more complicated now and will not be discussed here.

Our first corollary says that the MAP-GEXIT function can be computed by the DE analysis in high/low noise regimes. It also shows that the replica expressions computed at the appropriate fixed point are exact.

Corollary 1 (density evolution allows us to compute MAP). *Consider communication over channels $\mathcal{K}$. For ensembles* LDGM(Λ, P) *with high enough noise $\epsilon'_g < \epsilon < \epsilon_{\max}$ and* LDPC(Λ, P) *with low enough noise $0 < \epsilon < \epsilon'_p$ we have*

$$\lim_{n\to+\infty} g_n(\epsilon) = \lim_{d\to+\infty}\lim_{n\to+\infty} g_{n,d}^{BP}(\epsilon). \tag{17}$$

Here ϵ'_g and ϵ'_p depend only on $l_{\max}$,$k_{\max}$.

This result extends to the class of channels $\mathcal{K}$, those obtained previously on the BEC [1], [6]. In the case of LDPC ensembles with a vanishing GEXIT curve for $\epsilon \le \epsilon_*$, it is known that the result can be more easily obtained by physical degradation [7] or correlation inequalities [13], [14] for $\epsilon \le \epsilon_*$. However, there are ensembles with a GEXIT curve that is nontrivial all the way down to $\epsilon \to 0$ (for example, the Poisson LDPC

[3] See [19] for a nice discussion of this point related to the Griffith's singularity in the spin glass context.

ensemble) and for which the theorem is new. Note that it applies whether or not there is a phase transition (e.g., a jump discontinuity in the GEXIT curve), so it applies even in situations where the area theorem does not allow us to prove (17). The values obtained for $\epsilon'_{p,g}$ are worse than those $\epsilon_{p,g}$ obtained in Theorem 1. This is not surprising in view of the following remarks. It is expected (and for the BEC in some cases it is known) that the equality (17) is true as long as the noise parameter does not lie in a window around the phase transition threshold where this window is determined by an extended form of the BP-GEXIT curve (an S shaped curve). On the other hand inside the window, close to the phase transition threshold, it is known that (17) cannot hold. A look at the proof shows that the decay of correlations always implies (17) only if this decay is fast enough to beat the expansion of the graph: in other words if $\xi \ln(l_{\max} k_{\max}) \ll 1$. Our estimates allow us to control the growth of ξ^{-1} with respect to ϵ to show that such a regime exists. Therefore in a window close to the phase transition threshold, even if the correlations decay, $\xi \ln(l_{\max} k_{\max}) \ll 1$ cannot be valid.

Finally, concerning the exchange of limits $d, n \to +\infty$ for the BP algorithm, we prove what follows.

THEOREM 2 (exchange of limits). *Consider communication over the BSC. For* LDGM(Λ, P) *ensembles with bounded degrees with high enough noise* $\epsilon''_g < \epsilon < \epsilon_{\max}$, *depending only on* $l_{\max}, k_{\max}$, *we have*

$$\lim_{d\to+\infty} \lim_{n\to+\infty} g^{BP}_{n,d}(\epsilon) = \lim_{n\to+\infty} \limsup_{d\to+\infty} g^{BP}_{n,d}(\epsilon) = \lim_{n\to+\infty} \liminf_{d\to+\infty} g^{BP}_{n,d}(\epsilon). \tag{18}$$

The proof is a simple application of the decay of correlations. We present it only for the BSC, but it can also be extended to any convex combination of such channels and more generally as long as $c(l)$ has a bounded support that diminishes as the noise parameter increases. The cases of unbounded support (such as BIAWGNC) or of LDPC codes at low noise require more work and will not be discussed here. The present result complements the recent work [15], which concerns the bit-error-rate of LDPC codes for other message-passing decoders in the regime where the error rate vanishes.

3. LDGM codes: High noise. In this section we prove Theorem 1 and its corollary for LDGM codes. It is convenient to set $K = l_{\max} k_{\max}$.

Proof of Theorem 1 (LDGM). First we define the self-avoiding walks on which the cluster expansion is based. A self-avoiding walk w between two variable (information-bit) nodes a, b is a sequence of variable nodes (denoted $v_1, v_2, \dots, v_{l+1}$) and checks (denoted $c_1, c_2, \dots, c_l$), $v_1, c_1, v_2, c_2, \dots, c_l, v_{l+1}$ such that $v_1 = a$, $v_{l+1} = b$, $\{v_m, v_{m+1}\} \in \partial c_m$, and $v_m \neq v_n$, $c_m \neq c_n$ for $m \neq n$. We also say that two variable nodes a, b are connected if and only if there exists a self-avoiding walk from a to b. Thus on a self-avoiding walk we do not repeat variable and check nodes. From any general walk between a and b we can extract a self-avoiding walk w between a and b which has all its clauses belonging to the parent walk (this is done by chopping off all the loops of the general walk). The length $|w|$ of the walk is the number of variable nodes in it. If $a = b$, then the self-avoiding walk from a to b is the trivial walk a. We define the length of such walks to be zero. Let W_{ab} denote the set of all self-avoiding walks between variable nodes a, b and $W_{AB} = \cup_{a\in A, b\in B} W_{ab}$ (see Figure 1). Fix some number $H > 0$ (that will depend on ϵ later on). Denote by $\mathcal{B}$ the set of all code-bit nodes i (checks) such that $|l_i| > H$. We use the following (see Appendix A for the proof).

LEMMA 1. *Consider any LDGM code with bounded left and right degree. Consider two sets of information-bit nodes A, B with bounded support. We have*

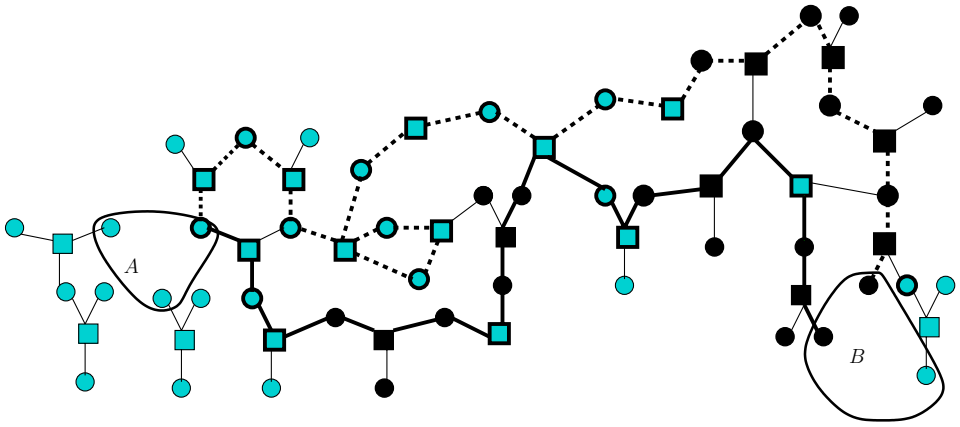

FIG. 1. *Each set A and B contains three variable nodes. The light squares denote the generator bits in the complement of $\mathcal{B}$, and the dark squares denote the generator bits in $\mathcal{B}$. The thick path is an example of a self-avoiding path between A and B which contributes to the upper bound. The dashed path is a non-self-avoiding path and does not contribute to the bound.*

$$\left|\left\langle \prod_{a\in A} u_a \prod_{b\in B} u_b \right\rangle - \left\langle \prod_{a\in A} u_a \right\rangle \left\langle \prod_{b\in B} u_b \right\rangle\right| \le 2 \sum_{w\in W_{AB}} \prod_{i\in w} \rho_i, \tag{19}$$

where $\rho_i = 1$ if $i \in \mathcal{B}$ and $\rho_i = e^{4|l_i|} - 1$ if $i \notin \mathcal{B}$.

The crucial feature of this lemma is that the ρ_i are independent random variables because the walks are self-avoiding. Consequently, averaging over the noise realization in (19),

$$\mathbb{E}_{l^n}\left|\left\langle \prod_{a\in A} u_a \prod_{b\in B} u_b \right\rangle - \left\langle \prod_{a\in A} u_a \right\rangle \left\langle \prod_{b\in B} u_b \right\rangle\right| \le 2 \sum_{w\in W_{AB}} \prod_{i\in w} \mathbb{E}[\rho_i]. \tag{20}$$

Now

$$\begin{aligned} \mathbb{E}[\rho_i] &\le \mathbb{E}[\rho_i | i \notin \mathcal{B}]\mathbb{P}(i \notin \mathcal{B}) + \mathbb{E}[\rho_i | i \in \mathcal{B}]\mathbb{P}(i \in \mathcal{B}) \\ &\le (e^{4H} - 1) + \mathbb{P}(|l| > H) = \delta(\epsilon, H). \end{aligned} \tag{21}$$

For our class of channels we can choose $H = H(\epsilon)$ such that $K\delta(\epsilon, H(\epsilon)) < 1$. We get

$$\begin{aligned} \mathbb{E}_{l^n}\left|\left\langle \prod_{a\in A} u_a \prod_{b\in B} u_b \right\rangle - \left\langle \prod_{a\in_A} u_a \right\rangle \left\langle \prod_{b\in_B} u_b \right\rangle\right| &\le 2 \sum_{w\in W_{AB}} \delta(\epsilon)^{|w|} \\ &\le 2|A||B| \sum_{d\ge \mathrm{dist}(A,B)} (K\delta(\epsilon, H(\epsilon)))^d \\ &\le \frac{2|A||B|}{1 - K\delta(\epsilon, H(\epsilon))} (K\delta(\epsilon))^{\mathrm{dist}(A,B)}. \end{aligned} \tag{22}$$

The second inequality is obtained by noticing that the number of self-avoiding random walks of length $|w|$ is certainly bounded by $K^{|w|}$. The factor $|A||B|$ accounts for the maximum possible number of initial and final vertices. The correlation decay of the theorem is in fact a special case of this last bound for the choice $A = \partial i$ and $B = \partial j$. ∎

We now look at GEXIT functions of the MAP and BP decoders. For LDGM codes we recall that the functional giving the MAP-GEXIT function in (10) is

$$\mathcal{G}(\langle x_i\rangle_0) = \frac{\Lambda'(1)}{P'(1)} \int dl_i \frac{dc(l_i)}{d\epsilon} \mathbb{E}_{l^{n\setminus i}} \ln\left\{\frac{1+\langle x_i\rangle_0 \tanh\, l_i}{1+\tanh\, l_i}\right\}. \tag{23}$$

The BP-GEXIT curve is given by the same functionals with $\langle x_i\rangle_0$ replaced by $\langle x_i\rangle_{0,d}^{BP}$. Consider $N_d(i)$ the neigborhood of node i, radius d an even integer (all the vertices at graph distance less or equal to d from i). As is well known for an ensemble LDGM(Λ, P) with bounded degrees, given d, if n is large enough, the probability that $N_d(i)$ is a tree is $1 - O(\frac{\gamma^d}{n})$ (where γ depends only on $l_{\max}, k_{\max}$; see, for example, [2] for an explicit proof). Thus when d is fixed and $n \to +\infty$, the computational tree $T_d(i)$ and the neighborhood $N_d(i)$ match with high probability. This implies that

$$\lim_{d\to+\infty} \lim_{n\to+\infty} g_{n,d}^{BP}(\epsilon) = \mathbb{E}_{\mathcal{C}}[\mathcal{G}(\langle x_i\rangle_{0,N_d(i)}) | N_d(i) \text{ is a tree}], \tag{24}$$

where $\langle x_i\rangle_{0,N_d(i)}$ is the Gibbs bracket associated to the subgraph $N_d(i)$. The right-hand side can be computed exactly by performing the statistical mechanical sums (see (6), (7), (8)) on a tree and yields the DE formulas

$$\lim_{d\to+\infty} \lim_{n\to+\infty} g_{n,d}^{BP}(\epsilon) = \lim_{d\to\infty} \frac{\Lambda'(1)}{P'(1)} \int dl \frac{dc(l)}{d\epsilon} \mathbb{E}_{\Delta^{(d)}} \ln\left\{\frac{1+\tanh\, \Delta^{(d)} \tanh\, l}{1+\tanh\, l}\right\}, \tag{25}$$

where both limits exist and

$$\tanh\, \Delta^{(d)} = \prod_{i=1}^{k} \tanh\, v_i^{(d)}. \tag{26}$$

The $v_i^{(d)}$ are i.i.d. random variables with distribution obtained from the iterative system of DE equations

$$\eta^{(d)}(v) = \sum_l \frac{l\Lambda_l}{\Lambda'(1)} \int \prod_{i=1}^{l-1} du_i \hat{\eta}^{(d)}(u_i) \delta\left(v - \sum_{i=1}^{l-1} u_i\right), \tag{27}$$

$$\hat{\eta}^{(d)}(u) = \sum_k \frac{kP_k}{P'(1)} \int dlc(l) \prod_{a=1}^{k-1} dv_a \eta^{(d-1)}(v_a) \delta\left(u - \tanh^{-1}\left(\tanh\, l \prod_{i=1}^{k-1} \tanh\, v_a\right)\right), \tag{28}$$

with the initial condition $\eta^{(0)}(v) = \delta(v)$. It is well known that these equations are an iterative version of the replica fixed point equation [23].

Proof of Corollary 1 (LDGM). Expanding the logarithm in (23) and using Nishimori identities as in [13], we obtain the expansion

$$\frac{\Lambda'(1)}{P'(1)} \sum_{p=1}^{+\infty} \frac{T_{2p}(\epsilon)}{2p(2p-1)} (\mathbb{E}_{\mathcal{C},l^{n\setminus i}}[\langle x_i\rangle_0^{2p}] - 1), \tag{29}$$

where we recall that

$$T_{2p}(\epsilon) = \frac{d}{d\epsilon} \int_{-\infty}^{+\infty} dlc(l)(\tanh\ l)^{2p}. \tag{30}$$

Note that in order to get the above expansion, it is important to use (23) as expressed here in terms of the *extrinsic* estimate. Obviously, the series is absolutely convergent, uniformly with respect to n, for the class of channels $\mathcal{K}$. Thus by dominated convergence, the proof will be complete if we show that

$$\lim_{n\to+\infty} \mathbb{E}_{\mathcal{C},l^{n\setminus i}}[\langle x_i\rangle_0^{2p}] = \lim_{d\to+\infty} \mathbb{E}_{\mathbf{\Delta}^{(d)}}[(\tanh\ \mathbf{\Delta}^{(d)})^{2p}]. \tag{31}$$

Indeed one can then compute the $n \to +\infty$ limit term by term in (29) and then resum the resulting series (which is again absolutely convergent, uniformly with respect to d) to obtain (25).

Let us show (31). As pointed out before for d fixed and $n \to +\infty$, $N_d(i)$ is a tree with high probability. Thus,

$$\lim_{n\to\infty} \mathbb{E}_{\mathcal{C},l^{n\setminus i}}[\langle x_i\rangle_0^{2p}] = \lim_{n\to\infty} \mathbb{E}_{\mathcal{C},l^{n\setminus i}}[\langle x_i\rangle_0^{2p} | N_d(i) \text{ is a tree}]. \tag{32}$$

Notice that all paths connecting the bit i with those outside $N_d(i)$ have a length at least equal to d, so because of Theorem 1 in the high noise regime x_i is very weakly correlated to the complement of $N_d(i)$. Therefore we may expect that

$$\lim_{d\to+\infty} \lim_{n\to+\infty} \mathbb{E}_{\mathcal{C},l^{n\setminus i}}[|\langle x_i\rangle_0^{2p} - \langle x_i\rangle_{0,N_d(i)}^{2p}| | N_d(i) \text{ tree}] = 0. \tag{33}$$

Assuming for a moment that this is true, we get from (32) that

$$\lim_{n\to\infty} \mathbb{E}_{\mathcal{C},l^{n\setminus i}}[\langle x_i\rangle_0^{2p}] = \lim_{d\to\infty} \lim_{n\to\infty} \mathbb{E}_{\mathcal{C},l^{n\setminus i}}[\langle x_i\rangle_{0,N_d(i)}^{2p} | N_d(i) \text{ is a tree}] \tag{34}$$

and, when $N_d(i)$ is a tree, the Gibbs average $\langle x_i\rangle_{0,N_d(i)}^{2p}$ is explicitly computable and that the right-hand side of (34) reduces to

$$\lim_{d\to+\infty} \mathbb{E}_{\mathbf{\Delta}^{(d)}}[(\tanh\ \mathbf{\Delta}^{(d)})^{2p}]. \tag{35}$$

This proves (31).

Our task is now to prove (33). Let $\mathring{N}_d(i)$ be the set of checks that are at distance d from i. We order the checks $\in \mathring{N}_d(i)$ in a given (arbitrary) way and call $\langle - \rangle_{0;\le k}$ the Gibbs average with $l_k = 0$ for the k first checks of $\mathring{N}_d(i)$ (and $l_i = 0$ for the root node). For the first one (call it 1) we use $e^{l_1 x_1} = \cosh\ l_1 + x_1 \sinh\ l_1$ to find

$$\langle x_i\rangle_0 = \langle x_i\rangle_{0;\le 1} + \frac{\tanh\ l_1(\langle x_i x_1\rangle_{0;\le 1} - \langle x_i\rangle_{0;\le 1}\langle x_1\rangle_{0;\le 1})}{1 + \langle x_1\rangle_{0;\le 1} \tanh\ l_1}. \tag{36}$$

Therefore

$$\begin{aligned} |\langle x_i\rangle_0^{2p} - \langle x_i\rangle_{0;\le 1}^{2p}| &\le 2p|\langle x_i\rangle_0 - \langle x_i\rangle_{0;\le 1}| \\ &\le 2pt_1|\langle x_i x_1\rangle_{0;\le 1} - \langle x_i\rangle_{0;\le 1}\langle x_1\rangle_{0;\le 1}| \end{aligned} \tag{37}$$

where

$$t_k = \frac{|\tanh\, l_k|}{1 - |\tanh\, l_k|}. \tag{38}$$

We can now take the second check of $\mathring{N}_d(i)$ (call it 2) and show

$$|\langle x_i\rangle_{0;\le 1}^{2p} - \langle x_i\rangle_{0;\le 2}^{2p}| \le 2pt_2|\langle x_i x_2\rangle_{0;\le 2} - \langle x_i\rangle_{0;\le 2}\langle x_2\rangle_{0;\le 2}|. \tag{39}$$

We can repeat this argument for all nodes of $\partial N_d(i)$ and use the triangle inequality to obtain

$$|\langle x_i\rangle_0^{2p} - \langle x_i\rangle_{0,N_d(i)}^{2p}| \le 2p \sum_{k\in \mathring{N}_d(i)} t_k|\langle x_i x_k\rangle_{0;\le k} - \langle x_i\rangle_{0;\le k}\langle x_k\rangle_{0;\le k}|. \tag{40}$$

Indeed the Gibbs average with all $l_k = 0$ for all $k \in \mathring{N}_d(i)$ is equal to $\langle x_i\rangle_{0,N_d(i)}$. Now using the bound (22) in the proof of Theorem 1 for $K\delta(\epsilon) < 1$, the last inequality implies

$$\mathbb{E}_{\mathcal{C},l^{n\backslash i}}[|\langle x_i\rangle_0^{2p} - \langle x_i\rangle_{0,N_d(i)}^{2p}||N_d(i)\text{ tree}] \le \frac{4pK^2\mathbb{E}[t]}{1 - K\delta(\epsilon)} K^d (K\delta(\epsilon))^d. \tag{41}$$

Note that for channels $\mathcal{K}$, for nonzero noise,

$$\mathbb{E}[t] = \mathbb{E}\left[\frac{|\tanh\, l|}{1 - |\tanh\, l|}\right] \le \mathbb{E}[e^{2|l|}] < \infty. \tag{42}$$

The right-hand side of (41) does not depend on n, so it is immediate that $\lim_{d\to+\infty}\lim_{n\to+\infty}$ vanishes as long as the noise is high enough such that $K^2\delta(\epsilon) < 1$. This proves (33) and the corollary. ∎

To conclude, let us remark that for the BIAWGNC, the GEXIT formulas simplify considerably, and there is a clear relationship to the magnetization,

$$\begin{aligned}\mathcal{G}(\langle x_i\rangle) &= \frac{1}{\epsilon^3}\frac{\Lambda'(1)}{P'(1)}(1 - \mathbb{E}_{l^n}[\langle x_i\rangle]) \\ &= \frac{1}{\epsilon^3}\frac{\Lambda'(1)}{P'(1)}(1 - \mathbb{E}_{l^n}[\tanh(l + \tanh^{-1}\langle x_i\rangle_0)])\end{aligned} \tag{43}$$

and

$$\lim_{d\to+\infty}\lim_{n\to+\infty} g_{n,d}^{BP}(\epsilon) = \frac{1}{\epsilon^3}\frac{\Lambda\,'(1)}{P\,'(1)}(1 - \mathbb{E}_{l,\Delta^{(d)}}[\tanh(l + \Delta^{(d)})]). \tag{44}$$

The proof of Corollary 1 for BIAWGNC can thus proceed without expansions and is slightly simpler. The main ideas can be found in [17], and we do not repeat them here. Note also that for the BEC there are similar simplifications that occur: this allows us to make a proof which avoids the second condition in the class of channels $\mathcal{K}$.

4. LDPC codes: Low noise. In this section we prove Theorem 1 and Corollary 1 for LDPC codes in a low noise regime. As explained in section 2, we first transform the problem to a dual one. The duality transformation reviewed here essentially is an application of Poisson's summation formula over commutative groups, and this point of view has been thoroughly discussed in the context of codes on graphs in [24]. Here we need to know how the correlations transform under the duality, a point that does not seem to appear in the related literature and that we review in the next paragraph.

4.1. Duality formulas for the correlations. Let C be a binary parity check code and $C^\perp$ its dual. We apply the Poisson summation formula

$$\sum_{x^n \in C} f(x^n) = \frac{1}{|C|} \sum_{\tau^n \in C^\perp} \hat{f}(\tau^n), \tag{45}$$

where the Fourier (or Hadamard) transform is

$$\hat{f}(\tau^n) = \sum_{x^n \in \{-1,+1\}^n} f(x^n) e^{i(\pi/4)\sum_{j=1}^n (1-\tau_j)(1-x_j)} \tag{46}$$

to the partition function Z of an LDPC code C. The dual code $C^\perp$ is an LDGM with codewords given by τ^n where

$$\tau_i = \prod_{a \in \partial i} u_a \tag{47}$$

and u_a are the m information bits. Noting that $Z = \sum_{x^n \in C} \prod_{i=1}^n e^{l_i x_i}$, (45) implies

$$Z = \frac{1}{|C^\perp|} \sum_{\tau \in C^\perp} \sum_{x^n \in \{-1,+1\}^n} \prod_{i=1}^n e^{l_i x_i + i(\pi/4)(1-\tau_i)(1-x_i)} \tag{48}$$

$$= \frac{1}{|C^\perp|} \sum_{\tau \in C^\perp} \prod_{i=1}^n (e^{l_i} + e^{-l_i} \tau_i) \tag{49}$$

$$= \frac{1}{|C^\perp|} e^{\sum_{i=1}^n l_i} \sum_{u^m \in \{-1,+1\}} \prod_{i=1}^n \Big(1 + e^{-2l_i} \prod_{a \in \partial i} u_a\Big). \tag{50}$$

To get the second equality above we exchanged the summations over x_i and the product over i and used the identity $e^{i(\pi/2)(1-\tau_i)} = \tau_i$ for $\tau_i = \pm 1$. To summarize, we have obtained

$$Z = \frac{1}{|C^\perp|} e^{\sum_{j=1}^n l_j} Z_\perp, \tag{51}$$

where

$$Z_\perp = \sum_{u^m \in \{-1,+1\}^m} \prod_{i=1}^n \left(1 + e^{-2l_i} \prod_{a \in \partial i} u_a\right). \tag{52}$$

This formula is known in coding theory as the "extended MacWilliams identity." Expression (52) formally looks like the partition function of an LDGM code with "channel half–log-likelihoods" g_i such that $\tanh g_i = e^{-2l_i}$. This is truly the case only for the BEC(ϵ) where $l_i = 0, +\infty$ and hence $g_i = +\infty, 0$, which still correspond to a BEC($1-\epsilon$). The logarithm of partition functions is related to the input-output entropy, and one recovers (taking the ϵ derivative) the well-known duality relation between EXIT functions of a code and its dual on the BEC [25]. For other channels, however, this is at best a formal (but still useful) analogy since the weights are negative for $l_i < 0$

(and g_i takes complex values). We introduce a bracket[4] $\langle - \rangle_\perp$,

$$\langle f \rangle_\perp = \frac{1}{Z_\perp} \sum_{u^m \in \{-1,+1\}^m} f(u^m) \prod_{i=1}^{n} \left(1 + e^{-2l_i} \prod_{a \in i} u_a \right). \tag{53}$$

The denominator may vanish, but it can be shown that when this happens, the numerator also does so in a way that ensures the finiteness of the ratio (this becomes clear in subsequent calculations). Taking logarithm of (51) and then the derivative with respect to l_i we find

$$\langle x_i \rangle = \frac{1}{\tanh 2l_i} - \frac{\langle \tau_i \rangle_\perp}{\sinh 2l_i}, \tag{54}$$

and differentiating once more with respect to l_j, $j \neq i$,

$$\langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle = \frac{\langle \tau_i \tau_j \rangle_\perp - \langle \tau_i \rangle_\perp \langle \tau_j \rangle_\perp}{\sinh 2l_i \sinh 2l_j}. \tag{55}$$

We stress that in (54), (55), τ_i and τ_j are given by products of information bits (47). The left-hand side of (54) is obviously bounded. It is less obvious to see this directly on the right-hand side, and here we just note that the pole at $l_i = 0$ is harmless since, for $l_i = 0$, the bracket has all its "weight" on configurations with $\tau_i = 1$. Similar remarks apply to (55). In any case, we will beat the poles by using the following trick. For any $0 < s < 1$ and $|a| \leq 1$ we have $|a| \leq |a|^s$, thus

$$\mathbb{E}_{l^n}[|\langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle|] \leq 2^{1-s} \mathbb{E}_{l^n}[|\langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle|^s]$$

and using (55) and Cauchy–Schwarz,

$$\mathbb{E}_{l^n}[|\langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle|] \leq 2^{1-s} \mathbb{E}[(\sinh 2l)^{-2s}] \mathbb{E}_{l^n}[|\langle \tau_i \tau_j \rangle_\perp - \langle \tau_i \rangle_\perp \langle \tau_j \rangle_\perp|^{2s}]^{1/2}. \tag{56}$$

The prefactor is always finite for $0 \leq s < \frac{1}{2}$ for our class of channels $\mathcal{K}$. For example, for the BIAWGNC we have

$$\mathbb{E}[(\sinh 2l)^{-2s}] \leq \frac{c}{|1-2s|} e^{-c'(s(1-2s))/\epsilon^2} \tag{57}$$

for purely numerical constants $c, c' > 0$, and for the BSC we have

$$\mathbb{E}[(\sinh 2l)^{-2s}] \leq \left(\frac{2\epsilon(1-\epsilon)}{1-2\epsilon} \right)^{2s}. \tag{58}$$

4.2. Decay of correlations for low noise. We will prove the decay of correlations by applying a high temperature cluster expansion technique to $\mathbb{E}_{l^n}[|\langle \tau_i \tau_j \rangle_\perp - \langle \tau_i \rangle_\perp \langle \tau_j \rangle_\perp|^{2s}]$. As explained in section 2, we need a technique that does not use the positivity of the Gibbs weights. In Appendix B we give a streamlined derivation of an adaptation of Berretti's expansion:

[4]In view of the comments above, this bracket is not a probabilistic expectation. For us the only important property that has to be retained is its linearity.

$$\langle \tau_i \tau_j \rangle_\perp - \langle \tau_i \rangle_\perp \langle \tau_j \rangle_\perp = \frac{1}{2} \sum_{\hat{X}} K_{i,j}(\hat{X}) \left(\frac{Z_\perp(\hat{X}^c)}{Z_\perp} \right)^2, \tag{59}$$

where

$$K_{i,j}(\hat{X}) \equiv \sum_{\substack{u_a^{(1)}, u_a^{(2)} \\ a \in \hat{X}}} \sum_{\substack{\Gamma \text{ compatible} \\ \text{with } \hat{X}}} (\tau_i^{(1)} - \tau_i^{(2)})(\tau_j^{(1)} - \tau_j^{(2)}) \prod_{k \in \Gamma} E_k \tag{60}$$

and

$$E_k = \tau_k^{(1)} e^{-2l_k} + \tau_k^{(2)} e^{-2l_k} + \tau_k^{(1)} \tau_k^{(2)} e^{-4l_k}. \tag{61}$$

Here $u_a^{(1)}$ and $u_a^{(2)}$ are two independent copies of the information bits (these are also known as real replicas) and $\tau_k^{(\alpha)} = \prod_{a \in k} u_a^{(\alpha)}$. To explain what are $\hat{X}$ and Γ we will refer to a-nodes (check nodes in the Tanner graph representing the LDPC code) and i-nodes (variable nodes in the Tanner graph representing the LDPC code). Given a subset S of nodes of the graph, let ∂S be the subset of neighboring nodes. In (59) the sum over $\hat{X}$ is carried over clusters of a-nodes such that "$\hat{X}$ is connected via hyperedges": this means that (a) $\hat{X} = \partial X$ for some connected subset X of i-nodes; (b) X is connected if any pair of i-nodes can be joined by a path, all of whose variable nodes lie in X; (c) $\hat{X}$ contains both ∂i and ∂j. In the sum (60) Γ is a set of i-nodes (all distinct). We say that "Γ is compatible with $\hat{X}$" if (i) $\partial\Gamma \cup \partial i \cup \partial j = \hat{X}$, (ii) $\partial\Gamma \cap \partial i \neq \phi$ and $\partial\Gamma \cap \partial j \neq \phi$, (iii) there is a walk connecting ∂i and ∂j such that all its variable nodes are in Γ. Finally,

$$Z_\perp(\hat{X}^c) = \sum_{\substack{u_a \\ a \in \hat{X}^c}} \prod_{\substack{\text{all } i \text{ such that} \\ \partial i \cap \hat{X} = \phi}} \left(1 + e^{-2l_i} \prod_{a \in i} u_a \right). \tag{62}$$

Figure 2 gives an example for all the sets appearing above.

We are now ready to prove the theorem on decay of correlations.

Proof of Theorem 1 (LDPC). Because of (56) it suffices to prove that $\mathbb{E}_{l^n}[|\langle \tau_i \tau_j \rangle_\perp - \langle \tau_i \rangle_\perp \langle \tau_j \rangle_\perp|^{2s}]$ decays.

The first step is to prove

$$\left| \frac{Z_\perp(\hat{X}^c)}{Z_\perp} \right| \leq 1. \tag{63}$$

This ratio cannot be estimated directly because the weights in $Z_\perp$ are not positive. However, we can use the duality transformation (51) backwards to get a new ratio of partition functions (corresponding to LDPC codes) with positive weights. More precisely, using the duality relation, we have

$$Z(\hat{X}^c) = \frac{1}{|C^\perp(\hat{X}^c)|} \left(\prod_{\substack{\text{all } k \text{ s.t.} \\ \partial k \cap \hat{X} = \phi}} e^{l_k} \right) \left(Z_\perp(\hat{X}^c) \right),$$

where

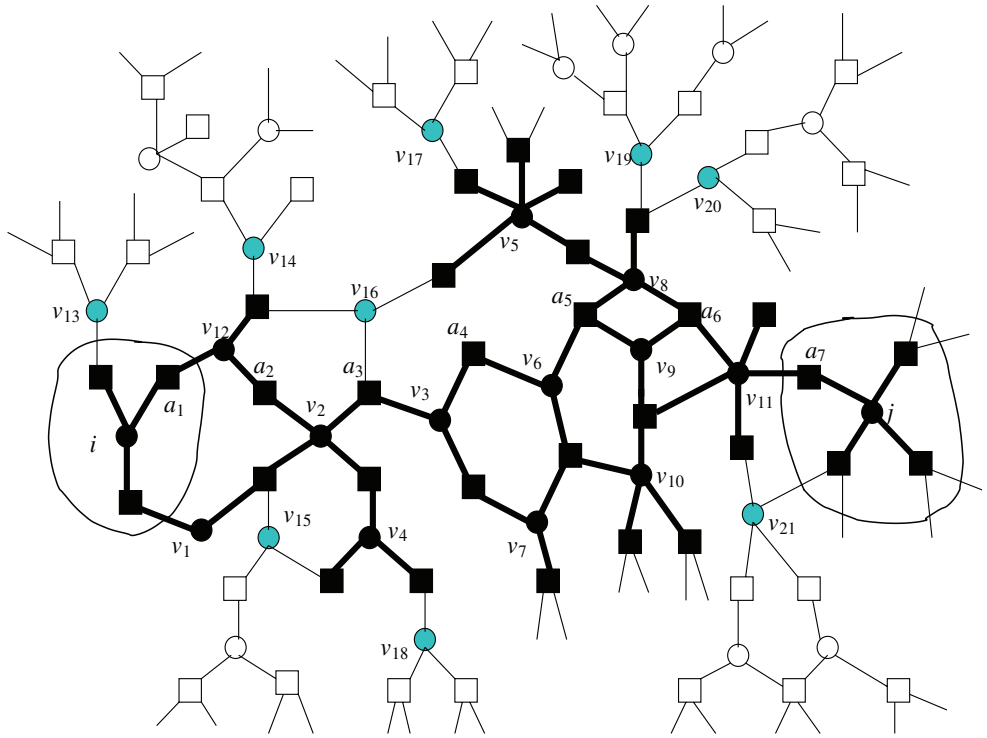

FIG. 2. *In this figure we explain the various sets appearing in the cluster expansion* (59). *The Tanner graph represents the LDPC code with variable nodes (i-nodes) denoted by circles and check nodes (a-nodes) denoted by squares. In this example the set* $\hat{X}$ *is the set of* dark *check nodes. Let us verify that this choice of* $\hat{X}$ *satisfies all our conditions. Let* $X = \{i, j, v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}, v_{12}\}$ *be a set of i-nodes (these are denoted by* dark *circles in the figure). It is easy to check that the set of neighbors of* X *is given by the* dark *check nodes, which is equal to the set* $\hat{X}$. *Hence* $\hat{X} = \partial X$, *and we satisfy condition* (a). *Secondly, any two variable nodes in* X *are connected by a path, all of whose variable nodes lie in* X; *thus condition* (b) *is met. Also notice that* $\hat{X}$ *contains both* ∂i *and* ∂j; *thus satisfying condition* (c). *Let us now show an example of the set* Γ *which appears in* (60). *Consider the set of distinct i-nodes,* $\Gamma = \{v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_{10}, v_{11}, v_{12}\}$. *In the figure, the set of i-nodes belonging to* Γ *have a dotted circle around them. Let us verify that* Γ *is compatible with* $\hat{X}$. *We meet the first condition* (i), *since it is easy to check that the set of a-nodes in the union* $\partial\Gamma \cup \partial i \cup \partial j$ *equals* $\hat{X}$. *To see that the condition* (ii) *is satisfied, consider the i-nodes* v_{12} *and* v_{11}. *Clearly,* $a_1 \in (\partial v_{12} \cap \partial i)$ *and* $a_7 \in (\partial v_{11} \cap \partial j)$. *The walk* $\{a_1 v_{12} a_2 v_2 a_3 v_3 a_4 v_6 a_5 v_8 a_6 v_{11} a_7\}$ *connects* ∂i *and* ∂j, *and all its i-nodes lie in* Γ; *hence the condition* (iii) *is also satisfied. Another choice for* Γ *would be the set* $\{v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}, v_{12}\}$. *In the definition of* $Z_\perp(\hat{X}^c), Z(\hat{X}^c)$ *the* light *variable nodes,* v_{13}, v_{14}, v_{15}, v_{16}, v_{17}, v_{18}, v_{19}, v_{20}, v_{21}, *are not present because they have a nonempty intersection with* $\hat{X}$ *(each of them has a* dark*a-node in their neighborhood). All check nodes which are* not dark *belong to* $\hat{X}^c$.

$$Z(\hat{X}^c) = \sum_{\substack{x_k \\ \partial k \cap \hat{X} = \phi}} \prod_{\substack{\text{all } k \text{ s.t.} \\ \partial k \cap \hat{X} = \phi}} e^{l_k x_k} \prod_{a \in \hat{X}^c} \frac{1}{2} \left(1 + \prod_{\substack{k \in a \text{ and} \\ \partial k \cap \hat{X} = \phi}} x_k \right). \tag{64}$$

This is the partition function corresponding to the subgraph induced by a-nodes of $\hat{X}^c$ and i-nodes such that $\partial i \cap \hat{X} = \phi$. Moreover $C^\perp(\hat{X}^c)$ is the dual of the code $C(\hat{X}^c)$.

Then combining with (51) we find

$$\frac{Z_\perp(\hat{X}^c)}{Z_\perp} = \left(\exp \sum_{\substack{\text{all } k \text{ s.t.} \\ \partial k \cap \hat{X} \neq \phi}} l_k\right)\left(\frac{|C^\perp(\hat{X}^c)|}{|C^\perp|}\right)\left(\frac{Z(\hat{X}^c)}{Z}\right). \tag{65}$$

To bound $\frac{Z_\perp(\hat{X}^c)}{Z_\perp}$ we first bound the ratio (of cardinality of the codes) $\frac{|C^\perp(\hat{X}^c)|}{|C^\perp|}$. The rank, $\mathfrak{L}$, of the parity-check matrix of code $C(\hat{X}^c)$, which is obtained by removing rows (checks) and columns (variables) from the parity-check matrix of the original LDPC code C, is smaller than the rank, $\mathfrak{T}$, of the parity-check matrix of C. Also set the block-length of the code $C(\hat{X}^c)$ to be n'. We have $|C| \cdot |C^\perp| = 2^n$ and $|C(\hat{X}^c)| \cdot |C^\perp(\hat{X}^c)| = 2^{n'}$. Dividing the two and using $|C(\hat{X}^c)| = 2^{n'-\mathfrak{L}}$ and $|C| = 2^{n-\mathfrak{T}}$ we get

$$\frac{|C^\perp(\hat{X}^c)|}{|C^\perp|} = \left(\frac{2^{n'}}{|C(\hat{X}^c)|}\right)\left(\frac{|C|}{2^n}\right) = 2^{n'-n}\left(\frac{2^{n-\mathfrak{T}}}{2^{n'-\mathfrak{L}}}\right) = 2^{\mathfrak{L}-\mathfrak{T}} \leq 1,$$

where in the last inequality we used that $\mathfrak{L} \leq \mathfrak{T}$.

Moreover we claim

$$\left(\exp \sum_{\substack{\text{all } k \text{ s.t.} \\ \partial k \cap \hat{X} \neq \phi}} l_k\right) Z(\hat{X}^c) \leq Z. \tag{66}$$

To see this one must recognize that the left-hand side of the inequality is the sum of terms of Z corresponding to x^n such that $x_k = +1$ for $\partial k \cap \hat{X} \neq \phi$ (and all other terms are ≥ 0). These remarks imply (63).

Using $|\sum_i a_i|^{2s} \leq \sum_i |a_i|^{2s}$ for $0 < 2s < 1$ and using (63) we find

$$\begin{aligned}\mathbb{E}_{l^n}[|\langle \tau_i \tau_j \rangle_\perp - \langle \tau_i \rangle_\perp \langle \tau_j \rangle_\perp|^{2s}] &= \frac{1}{2^{2s}} \mathbb{E}_{l^n} \left| \sum_{\hat{X}} K_{i,j}(\hat{X}) \left(\frac{Z_\perp(\hat{X}^c)}{Z_\perp}\right)^2 \right|^{2s} \\ &\leq \frac{1}{2^{2s}} \sum_{\hat{X}} \mathbb{E}_{l^n}[|K_{i,j}(\hat{X})|^{2s}]. \end{aligned} \tag{67}$$

Let us now provide an estimate on the right-hand side of the last inequality. Recall that

$$\begin{aligned} K_{i,j}(\hat{X}) &\equiv \sum_{\substack{u_a^{(1)}, u_a^{(2)} \\ a \in \hat{X}}} \sum_{\substack{\Gamma \text{ compatible} \\ \text{with } \hat{X}}} (\tau_i^{(1)} - \tau_i^{(2)})(\tau_j^{(1)} - \tau_j^{(2)}) \prod_{k \in \Gamma} E_k, \\ E_k &= \tau_k^{(1)} e^{-2l_k} + \tau_k^{(2)} e^{-2l_k} + \tau_k^{(1)} \tau_k^{(2)} e^{-4l_k}. \end{aligned} \tag{68}$$

To proceed further, first trivially bound the term $|(\tau_i^{(1)} - \tau_i^{(2)})(\tau_j^{(1)} - \tau_j^{(2)})|$ by 4 and $|E_k|$ by $|2e^{-2l_k} + e^{-4l_k}|$. Thus we obtain

$$|K_{i,j}(\hat{X})| \leq 4 \sum_{\substack{u_a^{(1)}, u_a^{(2)} \\ a \in \hat{X}}} \sum_{\substack{\Gamma \text{ compatible} \\ \text{with } \hat{X}}} \prod_{k \in \Gamma} |2e^{-2l_k} + e^{-4l_k}|$$
$$= 4^{|\hat{X}|+1} \sum_{\substack{\Gamma \text{ compatible} \\ \text{with } \hat{X}}} \prod_{k \in \Gamma} |2e^{-2l_k} + e^{-4l_k}|,$$

where to get the last equality we first notice that there is no more dependence on $u_a^{(1)}, u_a^{(2)}$ in the expression and we then use $\sum_{u_a^{(1)} u_a^{(2)}} = 4^{|\hat{X}|}$ for $a \in \hat{X}$.

Then we again use the identity $|2e^{-2l_k} + e^{-4l_k}|^{2s} \leq 2^{2s} e^{-4sl_k} + e^{-8sl_k}$ for $2s < 1$ to deduce

$$\mathbb{E}_{l^n}[|K_{i,j}(\hat{X})|^{2s}] \leq 4^{|\hat{X}|+1} \sum_{\substack{\Gamma \text{ compatible} \\ \text{with } \hat{X}}} (2^{2s}\mathbb{E}[e^{-4sl}] + \mathbb{E}[e^{-8sl}])^{|\Gamma|}$$
$$\leq 4^{|\hat{X}|+1} \sum_{\substack{\Gamma \text{ compatible} \\ \text{with } \hat{X}}} (\Delta(\epsilon))^{|\Gamma|}, \tag{69}$$

where in the first inequality we take the expectation inside the product because all the $k \in \Gamma$ are distinct and hence have independent l_k. We use the notation

$$\Delta(\epsilon) = 2^{2s}\mathbb{E}[e^{-4sl}] + \mathbb{E}[e^{-8sl}]. \tag{70}$$

It remains only to bound the number of Γ in the sum (69) and the cardinality of Γ which is equal to $|\Gamma|$. First we give a bound on $|\Gamma|$.

Remember that our final aim is to make the right-hand side of (69) small. Recall that the channel over which we are transmitting belongs to the class $\mathcal{K}$. From condition (3) of Definition 1 we can choose s and ϵ so that $\Delta(\epsilon)$ is less than one. To further estimate the bound in (69), we provide a *lower bound* on $|\Gamma|$. We do this as follows.

Since Γ is compatible with $\hat{X}$, we must have (from condition (i))

$$\partial\Gamma \cup \partial i \cup \partial j = \hat{X}.$$

As a consequence, we obtain

$$|\partial\Gamma| \geq |\hat{X}| - |\partial i| - |\partial j|. \tag{71}$$

Since the maximum degree of i-nodes is $\mathrm{l}_{\max}$, we have $|\partial\Gamma| \leq |\Gamma|\mathrm{l}_{\max}$, $|\partial i| \leq \mathrm{l}_{\max}$, and $|\partial j| \leq \mathrm{l}_{\max}$. Thus combining with (71), we get

$$|\Gamma| \geq (|\hat{X}| - 2\mathrm{l}_{\max})/\mathrm{l}_{\max}.$$

Let us now bound the number of Γ which are compatible with $\hat{X}$. Clearly, the maximum number of i-nodes which have an intersection with $\hat{X}$ is $|\hat{X}|\mathrm{r}_{\max}$. Thus there are at most $2^{|\hat{X}|\mathrm{r}_{\max}}$ possible choices for Γ. Combining all the above we obtain

$$\mathbb{E}_{l^n}[|K_{i,j}(\hat{X})|^{2s}] \leq 4^{(|\hat{X}|+1)} 2^{\mathrm{r}_{\max}|\hat{X}|} (\Delta(\epsilon))^{(|\hat{X}|-2\mathrm{l}_{\max})/\mathrm{l}_{\max}} \tag{72}$$

From (67) and (72) we get

$$\mathbb{E}_{l^n}[|\langle \tau_i\tau_j\rangle_\perp - \langle\tau_i\rangle_\perp\langle\tau_j\rangle_\perp|^{2s}] \le \frac{1}{2^{2s}}\sum_{\hat{X}} 4^{(|\hat{X}|+1)}2^{\mathrm{r}_{\max}|\hat{X}|}\Delta(\epsilon)^{(|\hat{X}|-2\mathrm{l}_{\max})/\mathrm{l}_{\max}}. \tag{73}$$

The clusters $\hat{X}$ connect ∂i and ∂j and thus have sizes $|\hat{X}| \ge \frac{1}{2}\mathrm{dist}(i,j)$.

Moreover we claim the following lemma.

LEMMA 2. *The number of clusters of a given size, $|\hat{X}|$, which satisfies conditions* (a), (b), (c), *is upper bounded by* $\mathrm{r}_{\max}|\hat{X}|(3^{K\mathrm{r}_{\max}|\hat{X}|})$, *where* $K = \mathrm{l}_{\max}\mathrm{r}_{\max}$ *is fixed for a given LDPC code.*

We prove this claim in Appendix B. Combining with Lemma 2 we can arrange the sum in (73) as

$$\begin{aligned}\mathbb{E}_{l^n}[|\langle \tau_i\tau_j\rangle_\perp - \langle\tau_i\rangle_\perp\langle\tau_j\rangle_\perp|^{2s}] &\le \frac{1}{2^{2s}}\sum_{|\hat{X}|\ge \mathrm{dist}(i,j)/2} \mathrm{r}_{\max}|\hat{X}|(3^{K\mathrm{r}_{\max}|\hat{X}|})4^{(|\hat{X}|+1)}2^{\mathrm{r}_{\max}|\hat{X}|}\\ &\qquad\times \Delta(\epsilon)^{(|\hat{X}|-2\mathrm{l}_{\max})/\mathrm{l}_{\max}}\\ &\le \frac{\mathrm{r}_{\max}(\Delta(\epsilon))^{-2}}{2^{2s}}\sum_{|\hat{X}|\ge \mathrm{dist}(i,j)/2} \exp\Big\{|\hat{X}|(\mathrm{r}_{\max}K\ \ln 3 + \ln 4\\ &\qquad + \mathrm{r}_{\max}\ \ln 2 + \frac{\ln(\Delta(\epsilon))}{\mathrm{l}_{\max}} + 1)\Big\}.\end{aligned} \tag{74}$$

For channels belonging to the class $\mathcal{K}$ we have, for s small enough, $\mathbb{E}[e^{-sl}] \to 0$ as $\epsilon \to 0$. As a consequence, we chose ϵ small enough to make $\Delta(\epsilon)$ small enough. Thus we can make the exponent negative, and we conclude the proof. ∎

4.3. Density evolution equals MAP for low noise. We recall that in the case of LDPC codes the functional giving the MAP-GEXIT function in (10) is

$$\mathcal{G}(\langle x_i\rangle_0) = \int dl_i \frac{dc(l_i)}{d\epsilon}\mathbb{E}_{l^{n\setminus i}}\ \ln\Big\{\frac{1+\langle x_i\rangle_0 \tanh\ l_i}{1+\tanh\ l_i}\Big\}. \tag{75}$$

The BP-GEXIT curve is given by the same functional with $\langle x_i\rangle_0$ replaced by the average on the computational tree $\langle x_i\rangle_{0,d}^{BP}$. As in section 3, we introduce $N_d(i)$, the neigborhood of node i, and radius d, an even integer. By the same arguments than in section 3 we have again

$$\lim_{d\to+\infty}\lim_{n\to+\infty} g_{n,d}^{BP}(\epsilon) = \mathbb{E}_C[\mathcal{G}(\langle x_i\rangle_{0,N_d(i)})|N_d(i) \text{ is a tree}], \tag{76}$$

where $\langle x_i\rangle_{0,N_d(i)}$ is the Gibbs bracket associated to the graph $N_d(i)$. It is important to note that for $N_d(i)$ a tree the set of leaves $\mathring{N}_d(i)$ are variable nodes and have "natural boundary conditions" as given by the channel outputs. The statistical mechanical sums on a tree yield the DE formula

$$\lim_{d\to+\infty}\lim_{n\to+\infty} g_{n,d}^{BP}(\epsilon) = \lim_{d\to\infty}\int dl \frac{dc(l)}{d\epsilon}\mathbb{E}_{\Lambda^{(d)}}\ \ln\Big\{\frac{1+\tanh\ \Lambda^{(d)}\tanh\ l}{1+\tanh\ l}\Big\}, \tag{77}$$

where both limits exist and

$$\Lambda^{(d)} = \sum_{a=1}^{k} w_a^{(d)}. \tag{78}$$

The $w_a^{(d)}$ are i.i.d. random variables with distribution obtained from the iterative system of DE equations

$$\zeta^{(d)}(w) = \sum_l \frac{l\Lambda_l}{\Lambda'(1)} \int \prod_{j=1}^{l-1} d\lambda_j \zeta^{(d)}(\lambda_j) \delta\Big(w - \tanh^{-1}\Big(\prod_{j=1}^{l-1} \tanh \lambda_j \Big)\Big),$$

$$\hat{\zeta}^{(d)}(\lambda) = \sum_k \frac{kP_k}{P'(1)} \int dl c(l) \prod_{a=1}^{k-1} dw_a \zeta^{(d-1)}(w_a) \delta\Big(\lambda - l - \sum_{a=1}^{k-1} w_a^{(d)} \Big),$$

with the initial condition $\eta^{(0)}(\lambda) = c(\lambda)$. As before, these equations are an iterative version of the replica fixed point equation [23].

Proof of Corollary 1 (LDPC). The first few steps are the same as in the proof for LDGM. First, we expand the logarithm in (75) and use Nishimori identities to obtain a series expansion like (29) (the prefactor $\frac{\Lambda'(1)}{P'(1)}$ is now absent). Second, we notice that since the resulting series expansion is uniformly absolutely convergent, it is enough to show that

$$\lim_{n\to+\infty} \mathbb{E}_{\mathcal{C},l^{n\setminus i}}[\langle x_i \rangle_0^{2p}] = \lim_{d\to+\infty} \mathbb{E}_d[(\tanh \Lambda^{(d)})^{2p}]. \tag{79}$$

Thirdly, as before, one argues that this follows from

$$\lim_{d\to+\infty} \lim_{n\to+\infty} \mathbb{E}_{C,l^{n\setminus i}}[|\langle x_i \rangle_0^{2p} - \langle x_i \rangle_{0,N_d(i)}^{2p}| | N_d(i) \text{ tree}] = 0, \tag{80}$$

and because of $|b^{2p} - a^{2p}| \le 2p|b-a|$ it is enough to show this for $2p$ replaced by 1. Unfortunately one cannot proceed as simply as in the LDGM case: (80) is a consequence of the next two auxiliary lemmas stated below.

Let $\langle - \rangle_{0;N_d(i)}^{\infty}$ be the bracket defined on the subgraph $N_d(i)$ with $l_k = +\infty$ for $k \in \partial N_d(i)$. This in fact is formally equivalent to fixing $x_k = +1$ boundary conditions on the leaves of the tree $k \in \mathring{N}_d(i)$. The first lemma says that the bit estimate can be computed locally.

LEMMA 3. *Under the same conditions as in Corollary 1,*

$$\lim_{d\to+\infty} \lim_{n\to+\infty} \mathbb{E}_{C,l^{n\setminus i}}[|\langle x_i \rangle_0 - \langle x_i \rangle_{0,N_d(i)}^{\infty}| | N_d(i) \text{ tree}] = 0. \tag{81}$$

The second lemma says that at low enough noise, free and +1 boundary conditions are equivalent

LEMMA 4. *Under the same conditions as in Corollary 1,*

$$\lim_{d\to+\infty} \lim_{n\to+\infty} \mathbb{E}_{C,l^{n\setminus i}}[\langle x_i \rangle_{0,N_d(i)}^{\infty} - \langle x_i \rangle_{0,N_d(i)} | | N_d(i) \text{ tree}] = 0. \tag{82}$$

We prove the first lemma. It will then be clear that the proof of the second one is essentially the same except that the original full graph is replaced by $N_d(i)$, and thus it will be spared.

Proof of Lemma 3. In (81) (and (82)) the root node i has $l_i = 0$, which turns out to be technically cumbersome because we really work in a low noise regime. For this reason we use

$$\langle x_i \rangle = \frac{\langle x_i \rangle_0 + \tanh\, l_i}{1 + \langle x_i \rangle_0 \tanh\, l_i}, \qquad \langle x_i \rangle^{\infty}_{N_d(i)} = \frac{\langle x_i \rangle^{\infty}_{0,N_d(i)} + \tanh\, l_i}{1 + \langle x_i \rangle^{\infty}_{0,N_d(i)} \tanh\, l_i} \tag{83}$$

to deduce

$$\langle x_i \rangle_0 - \langle x_i \rangle^{\infty}_{0,N_d(i)} = \frac{(1 - (\tanh\, l_i)^2)(\langle x_i \rangle - \langle x_i \rangle^{\infty}_{N_d(i)})}{(1 - \langle x_i \rangle \tanh\, l_i)(1 - \langle x_i \rangle^{\infty}_{N_d(i)} \tanh\, l_i)}. \tag{84}$$

This implies

$$|\langle x_i \rangle_0 - \langle x_i \rangle^{\infty}_{0;N_d(i)}| \le \frac{1 + |\tanh\, l_i|}{1 - |\tanh\, l_i|} |\langle x_i \rangle - \langle x_i \rangle^{\infty}_{N_d(i)}|, \tag{85}$$

and averaging over the noise and using Cauchy–Schwarz,

$$\begin{aligned} \mathbb{E}_{l^n \setminus i}[|\langle x_i \rangle_0 - \langle x_i \rangle^{\infty}_{0;N_d(i)}|] &\le 2\mathbb{E}[e^{4|l|}]^{1/2} \mathbb{E}_{l^n}[|\langle x_i \rangle - \langle x_i \rangle^{\infty}_{N_d(i)}|^2]^{1/2} \\ &\le 2\sqrt{2}\mathbb{E}[e^{8|l|}]^{1/2} \mathbb{E}_{l^n}[|\langle x_i \rangle - \langle x_i \rangle^{\infty}_{N_d(i)}|]^{1/2}. \end{aligned} \tag{86}$$

Let us now prove

$$\lim_{d \to +\infty} \lim_{n \to +\infty} \mathbb{E}_{C,l^n}[|\langle x_i \rangle - \langle x_i \rangle^{\infty}_{N_d(i)}| | N_d(i) \text{ tree}] = 0. \tag{87}$$

We order the variable nodes at the boundary $\mathring{N}_d(i)$ and consider the corresponding vector of log-likelihoods with components $\in \mathring{N}_d(i)$. If the first $k-1$ components of this vector is $l_1, \ldots, l_{k-1} = +\infty$, the kth component is l'_k, and the other ones are i.i.d. distributed as $c(l)$ (in other words they are "natural"), and we write $\langle - \rangle^{\infty}_{\le k-1}$. From the fundamental theorem of calculus, it is not difficult to see that

$$\begin{aligned} \langle x_i \rangle - \langle x_i \rangle^{\infty}_{N_d(i)} &= -\sum_{k \in \mathring{N}_d(i)} \int_{l_k}^{+\infty} dl'_k \frac{d}{dl'_k} \langle x_i \rangle^{\infty}_{\le k-1} \\ &= -\sum_{k \in \mathring{N}_d(i)} \int_{l_k}^{+\infty} dl'_k (\langle x_i x_k \rangle^{\infty}_{\le k-1} - \langle x_i \rangle^{\infty}_{\le k-1} \langle x_k \rangle^{\infty}_{\le k-1}). \end{aligned} \tag{88}$$

Using $|a| \le |a|^s$ for any $0 < s < 1$ and $|a| \le 1$ we get

$$|\langle x_i \rangle - \langle x_i \rangle^{\infty}_{N_d(i)}| \quad \le 2^{1-s} \sum_{k \in \mathring{N}_d(i)} \int_{l_k}^{+\infty} dl'_k |\langle x_i x_k \rangle^{\infty}_{\le k-1} - \langle x_i \rangle^{\infty}_{\le k-1} \langle x_k \rangle^{\infty}_{\le k-1}|^s. \tag{89}$$

Let $\langle - \rangle^{\infty,\perp}_{\le k-1}$ be the dual bracket (with the first k components of $\mathring{N}_d(i)$ $l_1 = \cdots = l_{k-1} = +\infty$ and the kth component equal to l'_k). Because of (55) we have

$$|\langle x_i\rangle - \langle x_i\rangle^{\infty}_{N_d(i)}|$$

$$\leq 2^{1-s}\sum_{k\in N_d(i)}\int_{l_k}^{+\infty} dl'_k \frac{|\langle \tau_i\tau_k\rangle^{\infty,\perp}_{\leq k-1} - \langle \tau_i\rangle^{\infty,\perp}_{\leq k-1}\langle \tau_k\rangle^{\infty,\perp}_{\leq k-1}|^s}{(\sinh 2l_i \sinh 2l_k')^{2s}}. \tag{90}$$

Note that the denominator in the integral is important to make the integral convergent for $l'_k \to \infty$. Moreover the singularity at $l_i, l'_k = 0$ is harmless as long as $2s < 1$. The next step is to use the cluster expansion in order to estimate >

$$\mathbb{E}_{l^n}\left[\int_{l_k}^{+\infty} dl'_k \frac{|\langle \tau_i\tau_k\rangle^{\infty,\perp}_{\leq k-1} - \langle \tau_i\rangle^{\infty,\perp}_{\leq k-1}\langle \tau_k\rangle^{\infty,\perp}_{\leq k-1}|^s}{(\sinh 2l_i \sinh 2l_k')^{2s}}\right]. \tag{91}$$

By following similar steps to those in the proof of Theorem 1 one obtains an upper bound similar to (72) except that the likelihoods of the end points are weighted differently, and therefore there are two factors of $\Delta(\epsilon)$ (see (70)) replaced by

$$\mathbb{E}\left[\frac{2^{2s}e^{-4sl} + e^{-8sl}}{(\sinh 2l)^{-2s}}\right] < \infty \quad \text{and} \quad \mathbb{E}\left[\int_l^{+\infty} dl\,' \frac{2^{2s}e^{-4sl\,'} + e^{-8sl\,'}}{(\sinh 2l\,')^{-2s}}\right] < \infty. \tag{92}$$

Finally we can average over the code ensemble conditional on the event that $N_d(i)$ is a tree. Since the clusters $\hat{X}$ that connect ∂i and ∂k, $k \in N_d(i)$, have size $|\hat{X}|k_{\max}$, we obtain the result as long as $\Delta(\epsilon)$ is small enough for ϵ small enough. ∎

5. Large block length versus large number of iterations. In the LDGM case we prove the exchange of limits $d, n \to +\infty$ for the BSC channel. As will become clear, one needs the decay of correlations (or covariance) of the Gibbs measure on the computational tree for $d \gg n$. Hence the likelihoods are not independent random variables: the proof of Theorem 1 still goes through in the case of the BSC. The only difference is that in Lemma 1 we can take $H > \frac{1}{2}\ln\frac{1-\epsilon}{\epsilon}$ such that $\mathcal{B} = \emptyset$ and $\rho_{\pi(j)} = e^{4|l_{\pi(j)}|} - 1 = \frac{4|1-2\epsilon|}{(1-|1-2\epsilon|)^2}$ for all $j \in T_d(i)$.

LEMMA 5 (decay of correlations for the BP decoder, LDGM on BSC). *Consider communication with a given LDGM code with block-length size n and bounded degrees $l_{\max}$, $k_{\max}$ over the* BSC(ϵ). *We can find $c > 0$, a small enough numerical constant such that for $l_{\max}k_{\max}|1-2\epsilon| < c$ we have, for any given realization of the channel outputs,*

$$|\langle x_i x_j\rangle^{BP}_d - \langle x_i\rangle^{BP}_d\langle x_j\rangle^{BP}_d| \leq c_1 e^{-c_2(\epsilon)\mathrm{dist}(i,j)}, \tag{93}$$

where i is the root of the computational tree, j is an arbitrary node, $c_1 > 0$ is a numerical constant, and $c_2(\epsilon) > 0$ depending only on ϵ, $l_{\max}$, $k_{\max}$. Moreover $c_2(\epsilon)$ increases like $\ln|1-2\epsilon|$ *as $\epsilon \to \frac{1}{2}$.*

Basically, this result is contained in [22] where it is obtained by Dobrushin's criterion. Note that it is valid for fixed noise realizations and not only on average. The unbounded case would require us to take averages, but then on the computational tree one has to control moments $\mathbb{E}[\rho_\pi(i)^m]$, and this requires more work. The following proof is a simple application of this lemma.

Proof of Theorem 2 (LDGM, BSC). We take for the number of iterations of the BP decoder $d \gg n$. On the computational tree $T_d(i)$ we consider the subtree of root i and depth $d' \ll n$. This subtree is a smaller computational tree $T_{d'}(i) \subset T_d(i)$ and $d' \ll n \ll d$. Let $\mathring{T}_{d'}(i)$ and the leaves k with $\mathrm{dist}(i,k) = d'$, and order them in an arbitrary way. Consider the Gibbs measure $\langle - \rangle^{BP}_{d;\leq k}$, where for the first k checks of $\mathring{T}_{d'}(i)$ we

set $l_{\pi(k)} = 0$ in (13). Proceeding as in section 3 we get

$$|\langle x_i\rangle_d^{BP} - \langle x_i\rangle_{d'}^{BP}| \leq \sum_{k\in \mathring{T}_{d'}(i)} t_{\pi(k)} |\langle x_i x_k\rangle_{d;\leq k}^{BP} - \langle x_i\rangle_{d;\leq k}^{BP}\langle x_k\rangle_{d;\leq k}^{BP}|. \tag{94}$$

For the BSC, $t_{\pi(k)} = \frac{|1-2\epsilon|}{1-|1-2\epsilon|}$. From Lemma 5 for $|1-2\epsilon|$ small enough (but independent of n, d)

$$|\langle x_i\rangle_d^{BP} - \langle x_i\rangle_{d'}^{BP}| = O(K^{d'} e^{-c_2(\epsilon)d'}). \tag{95}$$

In this equation $O(-)$ is uniformly bounded with respect to n and d (and the noise realizations of course). Recall the GEXIT function of the BP decoder

$$g_{n,d}(\epsilon) = \frac{\Lambda'(1)}{P'(1)} \int dl_i \frac{dc(l_i)}{d\epsilon} \mathbb{E}_{\mathcal{C},l^{n\setminus i}} \ln\left\{\frac{1+\langle x_i\rangle_{0,d}^{BP} \tanh\, l_i}{1+\tanh\, l_i}\right\}. \tag{96}$$

Since for $|1-2\epsilon| \ll 1$, $|\tanh\, l_i| = \frac{1}{2}|\ln\, \frac{1-\epsilon}{\epsilon}| \ll 1$, one can easily show

$$g_{n,d}(\epsilon) = g_{n,d'}(\epsilon) + O(K^{d'} e^{-c_2(\epsilon)d'}). \tag{97}$$

For example, one could proceed by expanding the ln in powers of $|\tanh\, l_i|$ and estimate the series term by term. Now since $O(-)$ is uniformly bounded with respect to n, d, (97) implies for d' fixed

$$\lim_{n\to+\infty} \liminf_{d\to\infty} g_{n,d}(\epsilon) = \lim_{n\to+\infty} g_{n,d'}(\epsilon) + O(K^{d'} e^{-c_2(\epsilon)d'}). \tag{98}$$

Now we take the limit $d' \to +\infty$,

$$\lim_{n\to+\infty} \liminf_{d\to\infty} g_{n,d}(\epsilon) = \lim_{d'\to+\infty} \lim_{n\to+\infty} g_{n,d'}(\epsilon). \tag{99}$$

A similar result with lim sup replacing lim inf is derived in the same way. ∎

6. Conclusion. Our main result is that the correlations decay in regimes of low noise for LDPC and high noise for LDGM codes. We stress here that this has been shown to hold for given codes, not for ensembles. An application of this result is, as we have shown, that the BP and MAP performance measures (given by GEXIT curves) coincide in these noise regimes. This also proves that the replica symmetric solutions of these graphical models are exact in the same noise regimes. Our emphasis here has not been on optimizing our estimates, but rather in setting up an analysis that is valid for a quite general class of channels. Indeed most rigorous results in this field are derived for the BEC channel, which is amenable to combinatorial techniques. In this paper we have shown that cluster expansion techniques of statistical mechanics are a valuable tool for the theory of error correcting codes on graphs, and we hope that more progress can be made along these lines.

We have not investigated the regimes of high noise for LDPC codes and low noise for LDGM codes. In the case of LDPC codes and high noise we are able to prove decay of correlations for ensembles that contain a sufficient fraction of degree one variable nodes. Indeed one can eliminate the degree one nodes and convert the problem to a new graphical model containing a mixture of hard parity-check constraints and soft LDGM-type weights. If the density of soft weights is high enough, the analysis of the present paper

can be extended (see [26] for a summary). Combining theses ideas with duality, one may also treat special ensembles of LDGM codes for low noise. This approach, however, is not entirely satisfactory, and it is not clear how to directly go about with cluster expansions in these regimes.

Let us mention a few directions that would be worth investigating along the lines of this work. Recently another decoding algorithm called "tree pruning"[5] [30], interpolating between BP and MAP and based on ideas of [31], has been proposed. This algorithm uses the "self-avoiding walk tree," and it would be interesting to investigate if there is a connection with the cluster expansions presented here which, at least for LDGM codes, are also based on self-avoiding paths. In [28] the authors derive a new type of expansion called "loop expansion" in an attempt to compute corrections to BP equations. The link to traditional cluster expansions is unclear to us. It would be interesting to develop rigorous methods to control the loop expansions. We would also like to point out the work [29] where a new derivation of the Gilbert–Varshamov bound is presented using the Mayer expansion for a hard-sphere systems.

We hope that the ideas and techniques investigated in the present work could have other applications in the broader context of random graphical models and specially constraint satisfaction problems. Let us mention that various forms of correlation decay have been investigated recently for the random K-satisfiability problem at low constraint density, by different methods [27]. This has allowed the authors to prove that the replica symmetric solution is exact at low constraint density. It remains to be seen if the cluster expansion techniques can extend such results to higher constraint densities and/or other constraint satisfaction problems.

Appendix A. Cluster expansions. In this appendix we explain the derivation of the two cluster expansions that we use. In the statistical mechanics literature these have been derived for spin systems with pair interactions on regular graphs. It turns out that they can be adapted to our setting. We try to give a self-contained but still reasonably short derivation here.

A.1. Cluster expansion for LDGM codes. Here we adapt the cluster expansion of Dreifus, Klein, and Perez in [20]. In the process we prove Lemma 1 stated in section 3. It will be very convenient to use the following compact notation:

$$\prod_{a\in X} u_a = u_X \quad \text{for any set } X \subset \{1, \dots, m\}. \tag{100}$$

In particular the code-bits $x_i = \prod_{a\in\partial i} u_a$ become $u_{\partial i}$, $i = 1, \dots, n$, and the correlation of Lemma 1 becomes

$$\langle u_A u_B\rangle - \langle u_A\rangle\langle u_B\rangle. \tag{101}$$

It is first necessary to rewrite the Gibbs measure (6) in a form such that the exponent is positive:

$$\frac{1}{Z}\prod_{i=1}^{m} e^{l_i x_i} = \frac{1}{Z'}\prod_{i=1}^{m} e^{l_i u_{\partial i} + |l_i|}, \tag{102}$$

where Z' is the appropriately modified partition function. We introduce the replicated measure, which is the product of two copies,

[5]Note that these authors use the same duality as used here.

$$\frac{1}{Z'^2}\prod_{i=1}^{m} e^{l_i(u_{\partial i}^{(1)}+u_{\partial i}^{(2)})+2|l_i|}. \tag{103}$$

Thus we now have two replicas of the information bits $u_1^{(1)}, \dots, u_m^{(1)}$ and $u_1^{(2)}, \dots, u_m^{(2)}$. The Gibbs bracket for the replicated measure is denoted by $\langle - \rangle_{12}$. It is easy to see that

$$\langle u_A u_B \rangle - \langle u_A \rangle \langle u_B \rangle = \frac{1}{2}\langle (u_A^{(1)} - u_A^{(2)})(u_B^{(1)} - u_B^{(2)}) \rangle_{12}. \tag{104}$$

Recall that $\mathcal{B} = \{i | |l_i| > H\}$ for some fixed number H, and set

$$e^{l_i(u_{\partial i}^{(1)}+u_{\partial i}^{(2)})+2|l_i|} - 1 = K_i. \tag{105}$$

It will be important to keep in mind later that $K_i \geq 0$. We have

$$\begin{aligned}
&\frac{1}{2}\langle (u_A^{(1)} - u_A^{(2)})(u_B^{(1)} - u_B^{(2)}) \rangle_{12} \\
&\qquad = \frac{1}{2Z'^2} \sum_{u^{(1)},u^{(2)}} f_A f_B \prod_{i\in\mathcal{B}} e^{l_i(u_{\partial i}^{(1)}+u_{\partial i}^{(2)})+2|l_i|} \prod_{i\in\mathcal{B}^c}(1+K_i) \\
&\qquad = \frac{1}{2Z'^2} \sum_{u^{(1)},u^{(2)}} f_A f_B \prod_{i\in\mathcal{B}} e^{l_i(u_{\partial i}^{(1)}+u_{\partial i}^{(2)})+2|l_i|} \sum_{G\subseteq\mathcal{B}^c} \prod_{i\in G} K_i \\
&\qquad = \frac{1}{2Z'^2} \sum_{G\subseteq\mathcal{B}^c} \sum_{u^{(1)},u^{(2)}} f_A f_B \prod_{i\in\mathcal{B}} e^{l_i(u_{\partial i}^{(1)}+u_{\partial i}^{(2)})+2|l_i|} \prod_{i\in G} K_i,
\end{aligned} \tag{106}$$

where $f_X = u_X^{(1)} - u_X^{(2)}$, $X = A, B$.

Take a term with given $G \subseteq \mathcal{B}^c$ in the last sum. We say that "G connects A and B" if and only if there exist a self-avoiding walk[6] w_{ab} with initial variable node $a \in A$, final variable node $b \in B$, and such that all check nodes of w_{ab} are in $G \cup \mathcal{B}$.[7] The crucial point is that if a set G does not connect A and B, then it gives a vanishing contribution to the sum. We defer the proof of this fact to the end of this section. For the moment let us show that it implies the bound in Lemma 1. The positivity of K_i implies

$$\begin{aligned}
&|\langle u_A u_B \rangle - \langle u_A \rangle \langle u_B \rangle| \\
&\qquad \leq \frac{2}{Z'^2} \sum_{\substack{G\subseteq\mathcal{B}^c \\ G \text{ connects} A \text{ and } B}} \sum_{u^{(1)},u^{(2)}} \prod_{i\in\mathcal{B}} e^{l_i(u_{\partial i}^{(1)}+u_{\partial i}^{(2)})+2|l_i|} \prod_{i\in G} K_i \\
&\qquad \leq \frac{2}{Z'^2} \sum_{w\in W_{AB}} \sum_{G'\subseteq\mathcal{B}^c\setminus w} \sum_{u^{(1)},u^{(2)}} \prod_{i\in\mathcal{B}} e^{l_i(u_{\partial i}^{(1)}+u_{\partial i}^{(2)})+2|l_i|} \prod_{i\in w\setminus\mathcal{B}} \mathfrak{h}_i \prod_{i\in G'} K_i.
\end{aligned} \tag{107}$$

In the second inequality we used $K_i \leq e^{4|l_i|} - 1 \equiv \mathfrak{h}_i$. Now resumming over $G' \subseteq \mathcal{B}^c \setminus w$ we obtain

[6] See section 3 for the definition of these walks.

[7] Note that it is really $G \cup \mathcal{B}$ that connects A and B. Since $\mathcal{B}$ is fixed, our definition is valid.

$$
\begin{aligned}
&|\langle u_A u_B\rangle - \langle u_A\rangle\langle u_B\rangle| \\
&\quad \le \frac{2}{Z'^2}\sum_{w\in W_{AB}}\prod_{i\in w\setminus\mathcal{B}}\mathfrak{h}_i\sum_{u^{(1)},u^{(2)}}\prod_{i\in\mathcal{B}} e^{l_i(u_{\partial i}^{(1)}+u_{\partial i}^{(2)})+2|h_i|}\prod_{i\in\mathcal{B}^c\setminus w}(1+K_i) \\
&\quad \le \frac{2}{Z'^2}\sum_{w\in W_{AB}}\prod_{i\in w\setminus\mathcal{B}}\mathfrak{h}_i\sum_{u^{(1)},u^{(2)}}\prod_{i\in\mathcal{B}} e^{l_i(u_{\partial i}^{(1)}+u_{\partial i}^{(2)})+2|h_i|}\prod_{i\in w\setminus\mathcal{B}}(1+K_i)\prod_{i\in\mathcal{B}^c\setminus w}(1+K_i) \\
&\quad = 2\sum_{w\in W_{AB}}\prod_{i\in w\setminus\mathcal{B}}\mathfrak{h}_i. \qquad (108)
\end{aligned}
$$

The second inequality follows by inserting extra terms $1+K_i\ge 1$ for $i\in w\setminus\mathcal{B}$ and the equality by reconstituting Z'^2 in the numerator. Now the last line is equal to

$$
2\sum_{w\in W_{AB}}\prod_{i\in w}\rho_i, \qquad \rho_i = 1, \qquad i\in\mathcal{B}, \quad \text{and} \quad \rho_i=\mathfrak{h}_i, \qquad i\notin\mathcal{B}. \tag{109}
$$

Hence the bound (19).

It remains to explain why, if G does not connect A and B, the G-term does not contribute to (106). Let $\partial G\cup\partial\mathcal{B}$ be the set of variable nodes connected to the check nodes $G\cup\mathcal{B}$. We define a partition $\partial G\cup\partial\mathcal{B}=V_A\cup V_C\cup V_B$ into three sets of variable nodes. V_A is the set of all variable nodes v such that there exist a self-avoiding walk w_{av} connecting some $a\in A$ to v and such that all ckeck nodes of w_{av} are in $G\cup\mathcal{B}$. V_B is similarly defined with B and $b\in B$ instead of A. Finally $V_C=(\partial G\cup\partial\mathcal{B})\setminus(V_A\cup V_B)$. By construction $V_C\cap V_A=V_C\cap V_B=\emptyset$. The point is that if G does not connect A and B, then $V_A\cap V_B=\emptyset$. Indeed, otherwise there would be a $u\in V_A\cap V_B$ with a walk w_{au} and a walk w_{ub}, both with all check nodes in $G\cup\mathcal{B}$, but this would mean that G connects A, and B through the walk $w_{au}\cup w_{ub}$. We also define three sets \of check nodes $C_A=(G\cup\mathcal{B})\cap\partial V_A$, $C_B=(G\cup\mathcal{B})\cap\partial V_B$, and $C_C=(G\cup\mathcal{B})\setminus(G_A\cup G_B)$. Again the three sets are disjoint when G does not connect A and B: indeed if there exists $c\in C_A\cap C_B$, then c belongs to both V_A and V_B, which we just argued is impossible. This situation is depicted in Figure 3.

Now we examine a term of (106) for a G that does not connect A and B. Expanding the product f_Af_B, using linearity of the bracket and symmetry under exchange of replicas $(1)\leftrightarrow(2)$, it is equal to the difference I–II, where

$$
\mathrm{I}=\frac{1}{2Z'^2}\sum_{u^{(1)},u^{(2)}}u_A^{(1)}u_B^{(1)}\prod_{i\in\mathcal{B}}e^{l_i(u_{\partial i}^{(1)}+u_{\partial i}^{(2)})+2|l_i|}\prod_{i\in G}K_i, \tag{110}
$$

$$
\mathrm{II}=\frac{1}{2Z'^2}\sum_{u^{(1)},u^{(2)}}u_A^{(1)}u_B^{(2)}\prod_{i\in\mathcal{B}}e^{l_i(u_{\partial i}^{(1)}+u_{\partial i}^{(2)})+2|l_i|}\prod_{i\in G}K_i. \tag{111}
$$

Because of the disjointness of the sets $V_{A,B,C}$ and $C_{A,B,C}$ (the areas enclosed in dotted lines; see Figure 3) one can, in I and II, factor the sums $\sum_{u^{(1)},u^{(2)}}$ in a product of three terms (in fact there is a fourth trivial term which is a power of two coming from the bits outside the dotted areas). Then by symmetry $1\leftrightarrow 2$ one recognizes that $\mathrm{I}=\mathrm{II}$. Thus I–II, and this proves that G does not contribute to (106) when it does not connect A and B.

A.2. Cluster expansion for LDPC codes. Here we adapt the Berretti cluster expansion to our setting. For more details we refer to [21], [19]. Consider the *replicated* partition function

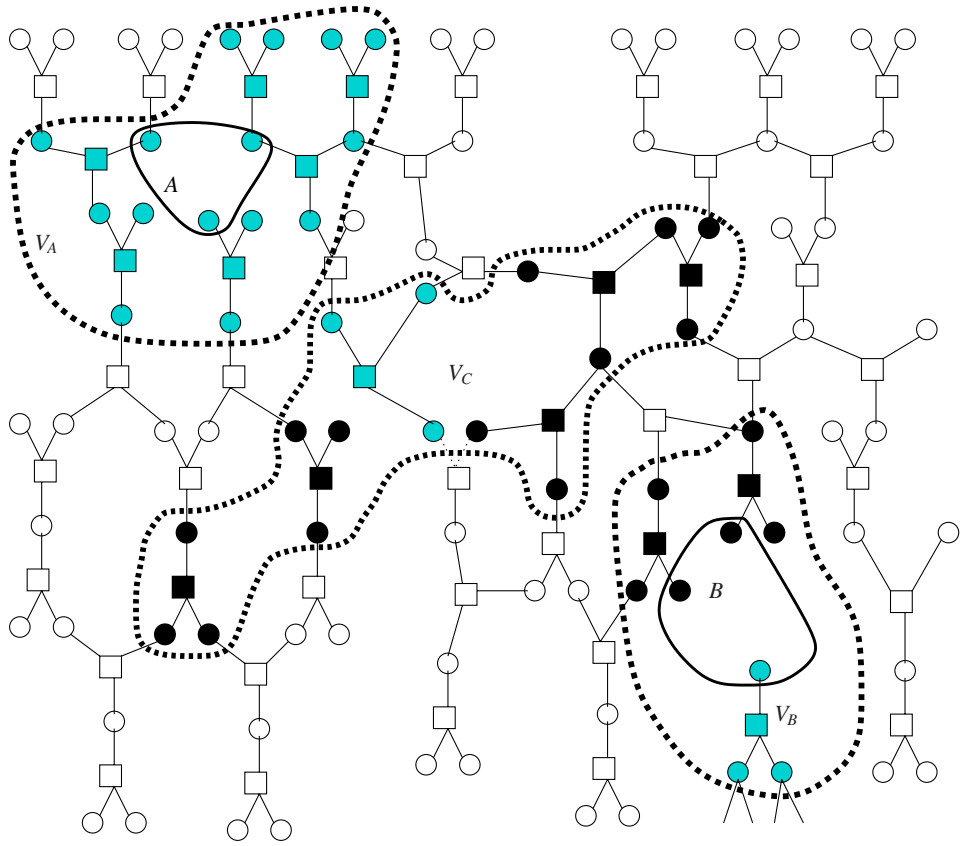

FIG. 3. *On the LDGM graph, $\mathcal{B}$ is depicted by the dark squares. The set $G \subseteq \mathcal{B}^c$ is depicted by the light squares. A and B both contain three nodes; there does not exist a self-avoiding walk that connects these two sets with all its check nodes in G. The sets of variable nodes V_A, V_B, and V_C are disjoint as well as the sets of check nodes G_A, G_B, and G_C: these sets are enclosed in the dotted areas.*

$$Z_\perp^2 = \sum_{u^{(1)},u^{(2)} \in \{-1,+1\}^m} \prod_{k=1}^{n} \left(1 + \tau_k^{(1)} e^{-2l_k}\right)\left(1 + \tau_k^{(2)} e^{-2l_k}\right). \tag{112}$$

Here $u^{(1)} = u_1^{(1)}, \ldots, u_m^{(1)}$, $u^{(2)} = u_1^{(2)}, \ldots, u_m^{(2)}$ are two replicas of the information bits and $\tau_k^{(1)} = \prod_{a \in k} u_a^{(1)}$, $\tau_k^{(2)} = \prod_{a \in k} u_a^{(2)}$. We have

$$\langle \tau_i \tau_j \rangle_\perp - \langle \tau_i \rangle_\perp \langle \tau_j \rangle_\perp = \frac{1}{2} \langle (\tau_i^{(1)} - \tau_i^{(2)})(\tau_j^{(1)} - \tau_j^{(2)}) \rangle_{\perp,12}, \tag{113}$$

where $\langle \cdot \rangle_{\perp,12}$ corresponds to the replicated system. We denote $f_i = \tau_i^{(1)} - \tau_i^{(2)}$, $f_j = \tau_j^{(1)} - \tau_j^{(2)}$. Then we have

$$\langle f_i f_j \rangle_{\perp,12} = \frac{1}{Z_\perp^2} \sum_{u^{(1)},u^{(2)}} f_i f_j \prod_k (1 + E_k), \tag{114}$$

where E_k is defined in (61). Expanding the product we get

$$\langle f_i f_j \rangle_{\perp,12} = \frac{1}{Z_\perp^2} \sum_{u^{(1)},u^{(2)}} f_i f_j \sum_{V \subset \mathfrak{V}} \prod_{k \in V} E_k$$

$$= \frac{1}{Z_\perp^2} \sum_{V \subset \mathfrak{V}} \sum_{u^{(1)},u^{(2)}} f_i f_j \prod_{k \in V} E_k, \tag{115}$$

where $\mathfrak{V}$ denotes the set of all variable nodes of the original Tanner graph for the LDPC code and V is any subset of distinct variable nodes. Suppose $V \subset \mathfrak{V}$ is such that *one cannot create a walk* (i.e., on the original Tanner graph of the LDPC code, a set of alternating variable and check nodes) connecting any check node in ∂i to any check node in ∂j and which has all its variable nodes contained entirely in V. Then we can partition V into three mutually disjoint sets of variable nodes, V_1, V_2, V_3 such that $V_1 \ni i$, $V_2 \ni j$, and $V_3 = V \setminus (V_1 \cup V_2)$. Note also that $\partial V_1, \partial V_2, \partial V_3$ are mutually disjoint; otherwise we can create a walk between ∂i and ∂j. Thus we can write

$$\sum_{u^{(1)},u^{(2)}} f_i f_j \prod_{k \in V} E_k = \sum_{\substack{u^{(1)},u^{(2)} \\ u_a^{(1)},u_a^{(2)} \in \partial V_1}} f_i \prod_{k \in V_1} E_k \sum_{\substack{u^{(1)},u^{(2)} \\ u_a^{(1)},u_a^{(2)} \in \partial V_2}} f_j \prod_{k \in V_2} E_k$$

$$\times \sum_{\substack{u^{(1)},u^{(2)} \\ u_a^{(1)},u_a^{(2)} \in \partial V_3}} \prod_{k \in V_3} E_k. \tag{116}$$

This implies that (116) vanishes. This is seen by using the antisymmetry of f_i (or f_j) and the symmetry of E_k, under the exchange $(1) \leftrightarrow (2)$. Thus only those V which contain a walk with all its variable nodes in V and which intersect both ∂i and ∂j contribute to the sum in (115).

For any given V (contributing to the sum) we construct the set of variable nodes Γ_V as follows. Γ_V is the union of all maximal connected clusters of distinct variable nodes in V such that each of those connected clusters intersects $\partial i \cup \partial j$. Let $\Gamma_V^c = V \setminus \Gamma_V$. Clearly, there exists such a set because we know that the walk which connects ∂i and ∂j is a subset of Γ_V. Let $\hat{X}_V = \partial \Gamma_V \cup \partial i \cup \partial j$ be a set of check nodes. It is not difficult to see that $\hat{X}_V$ satisfies all the requirements of the set $\hat{X}$ in the sum (59). Indeed, consider $X_V = \Gamma_V \cup i \cup j$. By construction $\partial X_V = \hat{X}_V$; any two variable nodes in X_V are connected by a walk with all its variable nodes in X_V; $\hat{X}_V$ contains both ∂i and ∂j. Also note that Γ_V is compatible with $\hat{X}_V$ as is required in the sum (60). Indeed, by construction $\partial \Gamma_V \cup \partial i \cup \partial j = \hat{X}_V$, $\partial \Gamma_V \cap \partial i \neq \phi$, and $\partial \Gamma_V \cap \partial j \neq \phi$; there exists a walk between ∂i and ∂j with all its variable nodes in Γ_V.

With this we can write

$$\langle f_i f_j \rangle_{\perp,12} = \frac{1}{Z_\perp^2} \sum_{V \subset \mathfrak{V}} \Bigg\{ \sum_{\substack{u_a^{(1)},u_a^{(2)} \\ a \in \partial \Gamma_V \cup \partial i \cup \partial j}} f_i f_j \prod_{k \in \Gamma_V} E_k \Bigg\} \Bigg\{ \sum_{\substack{u_a^{(1)},u_a^{(2)} \\ \text{remaining } a}} \prod_{k \in \Gamma_V^c} E_k \Bigg\}$$

$$= \frac{1}{Z_\perp^2} \sum_{\hat{X}} \sum_{\substack{V \subset \mathfrak{V}: \\ \hat{X}_V = \hat{X}}} \Bigg\{ \sum_{\substack{u_a^{(1)},u_a^{(2)} \\ a \in \hat{X}}} f_i f_j \prod_{k \in \Gamma_V} E_k \Bigg\} \Bigg\{ \sum_{\substack{u_a^{(1)},u_a^{(2)} \\ \text{remaining } a}} \prod_{k \in \Gamma_V^c} E_k \Bigg\}. \tag{117}$$

Now we resum over the sets V such that $\hat{X}_V = \hat{X}$. These consist of Γ compatible with $\hat{X}$ and the rest $\mathcal{G}$ which does not intersect $\hat{X}$. So

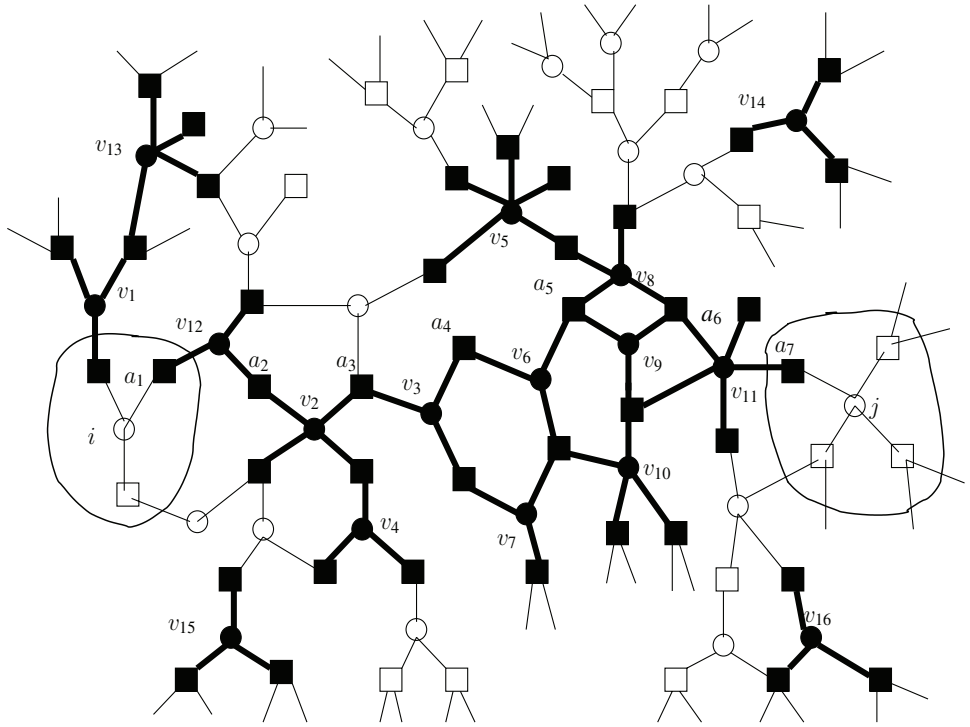

FIG. 4. *The* dark *variable nodes form the set* $V = \{v_1, \ldots, v_{16}\}$. *The walk* $a_1 v_{12} a_2 v_2 a_3 v_3 a_4 v_6 a_5 v_9 a_6 v_{11} a_7$ *connects* ∂i *to* ∂j, *and hence this* V *has a nonvanishing contribution.* $\Gamma_V = \{v_1, \ldots, v_{12}\}$ *is union of the two maximal connected clusters* $\{v_1, v_{13}\}$ *and* $\{v_{12}, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}\}$, *which has intersection with* $\partial i \cup \partial j$. $\Gamma_V^c = \{v_{14}, v_{15}, v_{16}\}$.

$$\langle f_i f_j \rangle_{\perp,12} = \frac{1}{Z_\perp^2} \sum_{\hat{X}} \Bigg\{ \sum_{\substack{u_a^{(1)}, u_a^{(2)} \\ a \in \hat{X}}} \sum_{\substack{\Gamma \text{ compatible} \\ \text{with } \hat{X}}} f_i f_j \prod_{k \in \Gamma} E_k \Bigg\} \Bigg\{ \sum_{\substack{u_a^{(1)}, u_a^{(2)} \\ a \in \hat{X}^c}} \sum_{\substack{\mathcal{G} \subset \mathfrak{B} \\ \partial \mathcal{G} \cap \hat{X} = \phi}} \prod_{k \in \mathcal{G}} E_k \Bigg\}$$

$$= \frac{1}{Z_\perp^2} \sum_{\hat{X}} \Bigg\{ \sum_{\substack{u_a^{(1)}, u_a^{(2)} \\ a \in \hat{X}}} \sum_{\substack{\Gamma \text{ compatible} \\ \text{with } \hat{X}}} f_i f_j \prod_{k \in \Gamma} E_k \Bigg\} \Bigg\{ \sum_{\substack{u_a^{(1)}, u_a^{(2)} \\ a \in \hat{X}^c}} \prod_{\substack{\text{all } k \text{ s.t.} \\ \partial k \cap \hat{X} = \phi}} (1 + E_k) \Bigg\}. \tag{118}$$

The last bracket is equal to (62), and we recognize Berretti's expansion. Figure 4 shows a sample set V and Γ_V which give a nonvanishing contribution.

Appendix B. Proof of Lemma 2. Before we commence, let us recall that a-nodes and i-nodes are the check nodes and variable nodes, respectively, of the Tanner graph of the LDPC code. Recall that $\mathrm{l}_{\max}$ denotes the maximum degree of i-nodes and $\mathrm{r}_{\max}$ denotes the maximum degree of a-nodes.

Let L be a positive integer. Our aim is to count the number of $\hat{X}$ such that $|\hat{X}| = L$ and the conditions (a), (b), and (c) given in section 4.2 are satisfied. Instead, we will count the number of $\hat{X}$ such that $|\hat{X}| = L$ and which satisfy conditions (a) and (b). But we require only that it contains ∂i. It may or may not contain ∂j. Clearly, this will be an upper bound to our required number.

We know that the required $\hat{X}$ corresponds to some set of i-nodes X such that $\partial X = \hat{X}$ (condition (a) in section 4.2). Also X is connected; i.e., there is a walk between any two i-nodes contained in X. Note that X may or may not contain the i-node i. Let us modify this X to include the i-node (if it is not already present). Let us denote this set by $\mathring{X}$. We point out two properties of the set $\mathring{X}$: (i) It is connected. Indeed, any two i-nodes, neither of which is i, are connected (follows from previous connectivity in X). Let $\partial^2 i$ denote the neighborhood of the set ∂i. It is clear that $\partial^2 i$ is a set of i-nodes. Since $\hat{X}$ includes ∂i, it must be that X contains at least one i-node which belongs to $\partial^2 i$. But i is already present in $\partial^2 i$. Consequently, i is connected to any other i-node in $\mathring{X}$. (ii) The size of $\mathring{X}$ is bounded as $\frac{L}{\mathrm{l}_{\max}} \leq |\mathring{X}| \leq L\mathrm{r}_{\max}$.

Thus for each $\hat{X}$ we have constructed an $\mathring{X}$ such that $\partial\mathring{X} = \hat{X}$, $i \in \mathring{X}$, $\mathring{X}$ is connected, and $|\mathring{X}| \in [\frac{L}{\mathrm{l}_{\max}}, L\mathrm{r}_{\max}]$. Suppose that $\hat{X}_1 \neq \hat{X}_2$ are two sets with size equal to L and both satisfy our requirements. Then it is clear that the corresponding $\mathring{X}_1$ and $\mathring{X}_2$ are also different. As a result, counting such $\mathring{X}$ will provide us with an upper bound on the number of required $\hat{X}$ of given size L.

Without loss of generality we consider the following graph, $\mathcal{G}$, to count the number of $\mathring{X}$. The nodes in $\mathcal{G}$ consists of the i-nodes in the original graph. We attach an edge between two nodes in $\mathcal{G}$ if there is an a-node containing the corresponding i-nodes in the original graph.

We simplify our analysis as follows. Instead of counting on $\mathcal{G}$, we perform our counting on the *computation tree*, denoted by $\mathcal{C}_G$ rooted at the node i. It is not hard to see that by counting $\mathring{X}$ on $\mathcal{C}_G$ we will overestimate the same on $\mathcal{G}$. We further simplify by considering the degree of the root node in $\mathcal{C}_G$ to be K and the degree of every other node to be $K+1$. Denote this last graph by $\mathcal{T}_G$.

With the above simplifications, let us now count the number of sets $\mathring{X}$ on the graph $\mathcal{T}_G$. We first provide a crude upper bound which is sufficient for the purpose.

We will now provide a *fingerprint* for any set $\mathring{X}$. Given a set $\mathring{X}$, generate the following sequence of numbers. Start from the root node and number every edge coming out of the root node as follows. Mark an edge e coming out of the root node by 0 if the node (other than the root node) is not contained in the set $\mathring{X}$. If the node is contained in $\mathring{X}$ but not further extended (by extended we mean we include any of the node below the current node in the set $\mathring{X}$), then mark the edge by 1. If the node is further extended below, then mark the edge by 2.

Consider the following generation of the fingerprint of the set $\mathring{X}$. Perform breadth-first-search (BFS)-type labeling of edges according to previous rules. First consider the K-tuple ($\in \{0,1,2\}$) corresponding to the edges emanating from the root node. Observe that the set $\mathring{X}$ grows only along edges labeled as 2. Thus write down the K-tuple corresponding to the end-node of all edges marked as 2. Proceed to build up the sequence of numbers in a BFS manner. At the end we have a sequence of K-tuples each containing either 0, 1, or 2. It is easy to see that there is a one-to-one correspondence between the fingerprint and the set $\mathring{X}$. Indeed, from a given finger-print sequence one can uniquely generate a set $\mathring{X}$ in a BFS manner. Note that we have written down a K-tuple only when we have added a node to the set $\mathring{X}$. Clearly the number of K-tuples in the fingerprint is upper bounded by L. As a result, the number of possible sets, $\mathring{X}$, is upper bounded by $(3^K)^{|\mathring{X}|}$.

Putting everything together we get that the number of sets $\hat{X}$, with a given size L, is upper bounded by

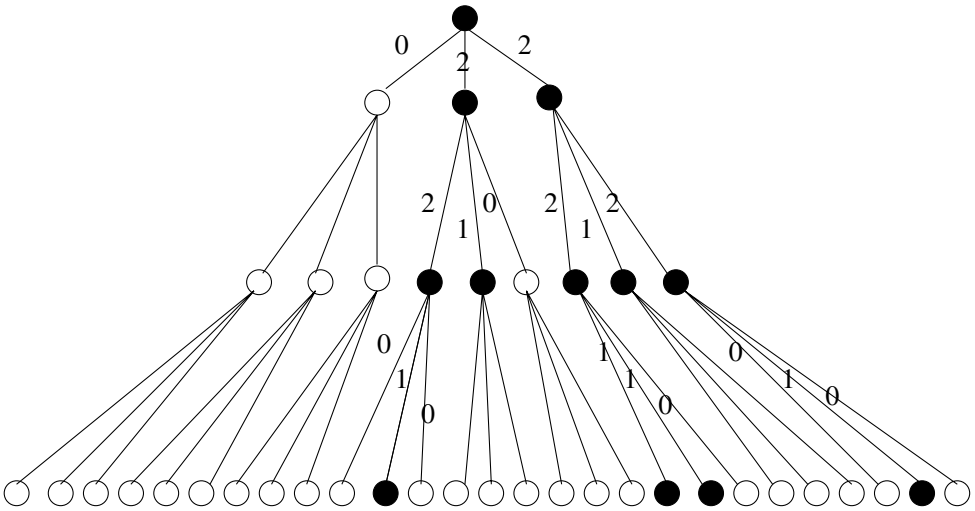

FIG. 5. *The fingerprint corresponding to the set* $\mathring{X}$ *shown by dark circles is given by* $F = \{(0,2,2)(2,1,0)(2,1,2)(0,1,0)(1,1,0)(0,1,0)\}$. *It is easy to see that given the fingerprint* F, *we can reconstruct the set* $\mathring{X}$ *uniquely in a BFS manner.*

$$\sum_{|\mathring{X}|=\frac{L}{\mathrm{r}_{\max}}}^{\mathrm{r}_{\max}L} (3^{K|\mathring{X}|}) \leq L\mathrm{r}_{\max}(3^{K\mathrm{r}_{\max}L}),$$

proving the lemma. Figure 5 illustrates a BFS labeling. □

REFERENCES

[1] T. RICHARDSON AND R. URBANKE, *Modern Coding Theory*, Cambridge University Press, London, 2008.

[2] T. RICHARDSON AND R. URBANKE, *The capacity of LDPC codes under message-passing decoding*, IEEE Trans. Inform. Theory, 47 (2001), pp. 638–656.

[3] A. MONTANARI, *Tight bounds for LDPC and LDGM codes under MAP decoding*, IEEE Trans. Inform. Theory, 51 (2005), pp. 3221–3246.

[4] S. KUDEKAR AND N. MACRIS, *Sharp bounds for optimal decoding of low-density-parity-check codes*, IEEE Trans. Inform. Theory, 53 (2007), pp. 2365–2375.

[5] F. GUERRA AND F. TONINELLI, *Quadratic replica coupling in the Sherrington-Kirkpatrick mean field spin glass model*, J. Math. Phys. 43 (2002), pp. 3704–3717.

[6] C. MEASSON, A. MONTANARI, AND R. URBANKE, *Maxwell construction: The hidden bridge between iterative and maximum a posteriori decoding*, IEEE Trans. Inform. Theory, 54 (2008), pp. 5277–5307.

[7] C. MEASSON, A. MONTANARI, T. RIDCHARSON, AND R. URBANKE, *The generalized area theorem and some of its consequences*, IEEE Trans. Inform. Theory, 55 (2009), pp. 4793–4821.

[8] S. FRANZ AND M. LEONE, *Replica bounds for optimization problems and diluted spin systems,* J. Stat. Phys., 111 (2003), pp. 535–564.

[9] M. TALAGRAND, *Spin Glasses: A Challenge for Mathematicians*, Spinger-Verlag, Berlin, 2003.

[10] S. KUDEKAR, S. KORADA, AND N. MACRIS, *Exact solution for the conditional entropy of Poissonian LDPC codes over the binary erasure channel*, in Proceedings of the IEEE International Symposium on Information Theory, Nice, France, 2007, pp. 1016–1021.

[11] C. MÉASSON, A. MONTANARI, AND R. URBANKE, *Asymptotic rate versus design rate*, in Proceedings of the IEEE International Symposium on Information Theory, Nice, France, 2007, pp. 1541–1545.

[12] H. O. GEORGII, *Gibbs Measures and Phase Transitions*, de Gruyter Stud. Math. 9, de Gruyter, Berlin, 1988.

[13] N. Macris, *Sharp bounds on generalized EXIT functions*, IEEE Trans. Inform. Theory, 53 (2007), pp. 2365–2375.

[14] N. Macris, *Griffith-Kelly-Sherman correlation inequalities: A useful tool in the theory of error correcting codes*, IEEE Trans. Inform. Theory, 53 (2003), pp. 664–683.

[15] S. Korada and R. Urbanke, *Exchange of limits: Why iterative decoding works*, IEEE Trans. Inform. Theory, submitted.

00 D. Brydges, *A short course on cluster expansions*, Course 3, in Critical Phenomena, Random Systems, Gauge Theories, in Course Notes of Proceedings of the Les Houches Summer School of Theoretical Physics, Session XLIII, Part 1, 1984, L'Ecole de Physique des Houches, pp. 129–183.

[17] S. Kudekar and N. Macris, *Proof of replica formulas in the high noise regime for communication using LDGM codes*, in Proceedings of the IEEE Information Theory Workshop, Porto, 2008, pp. 416–4120.

[18] S. Kudekar and N. Macris, *Decay of correlations in low density parity check codes: Low noise regime*, in *Proceedings of the IEEE International Symposium on Information Theory*, Seoul 2009.

[19] J. Fröhlich, *Mathematical aspects of disordered systems*, Course 9, in Critical Phenomena, Random Systems, Gauge Theories, in Course Notes of Proceedings of the Les Houches Summer School of Theoretical Physics, Session XLIII, Part 1, 1984, L'Ecole de Physique des Houches.

[20] H. Dreifus, A. Klein, and J. Perez, *Taming Griffiths' singularities: Infinite differentiability of quenched correlation functions*, Commun. Math. Phys., 170 (1995), pp. 21–39.

[21] A. Berretti, *Some properties of random Ising models*, J. Stat. Phys., 38 (1985), pp. 483–496.

[22] S. Tatikonda and M. I. Jordan, *Loopy belief propagation and Gibbs measures*, in Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence, Edmonton, Canada, 2002, Morgan Kaufmann, San Francisco.

[23] T. Murayama, Y. Kabashima, D. Saad, and R. Vicente, *The statistical physics of regular low-density parity-check error-correcting codes*, Phys. Rev. E, 62 (2000), pp. 1577–1591.

[24] D. Forney, *Codes on graphs: Normal realizations*, IEEE Trans. Inform. Theory, 47 (2001), pp. 520–548.

[25] A. Ashikmin, G. Kramer, and S. ten Brink, *Extrinsic information transfer functions: Model and erasure channel property*, IEEE Trans. Inform. Theory, 50 (2004), pp. 2657–2673.

[26] S. Kudekar and N. Macris, *Decay of correlations: An application to low density parity check codes*, in Proceedings of the 2008 IEEE 5th International Symposium on Turbo Codes and Related Topics, 2008, Lausanne, France, pp. 13–18.

[27] A. Montanari and D. Shah, *Counting good truth assignmants for random satisfiability formulae*, in Proceedings of the ACM-SIAM Symposium of Discrete Algorithms, New Orleans, 2007.

[28] M. Chertkov and V. Chernyak, *Loop series for discrete statistical models on graphs*, J. Stat. Mech. Theory Exp. (2006), P06009.

[29] A. Procacci and B. Scoppola, *Statistical mechanics approach to coding theory*, J. Stat. Phys., 96 (1999), pp. 907–912.

[30] Y. Lu, C. Measson, and A. Montanari, in *TP decoding*, in Proceedings of the 45th Annual Allerton Conference on Communication, Control and Computing, University of Illinois, 2007.

[31] D. Weitz, *Counting independent sets up to the threshold*, in Proceedings of the 38th ACM Symposium on Theory of Computing, Seattle, 2006.